



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost



MINISTERSTVO VNITRA
ČESKÉ REPUBLIKY

Dokumentace navržených mechanismů kontroly kvality datových sad katalogizovaných v Národním katalogu otevřených dat

Vytvořeno v rámci projektu

Implementace strategií v oblasti otevřených dat II
CZ.03.4.74/0.0/0.0/15_025/0004172

Klíčová aktivita: 02 Rozvoj Národního katalogu otevřených dat

Indikátor: 6 05 00 Počet napsaných a zveřejněných analytických a strategických dokumentů (vč. evaluačních)



Verze výstupu: 02

Verze k 17.4.2018

Popis výstupu

Tento výstup obsahuje úvod do problematiky datové kvality v kontextu otevřených dat a jejich katalogizace v Národním katalogu otevřených dat (NKOD). Jsou zde popsány jednotlivé techniky měření datové kvality pomocí záznamů v NKOD a jsou navrženy postupy pro její zlepšení.

Obsah

Popis výstupu	3
Pojmy	4
Kvalita datových sad katalogizovaných v NKOD	4
Obsah záznamu o datové sadě v NKOD	5
Položky popisující datovou sadu	5
Položky popisující distribuci datové sady	6
Cíle kontroly kvality datových sad	6
Kontrola metadat v NKOD	8
Pipeline v nástroji LinkedPipes ETL	8
Q1: Počet distribucí bez specifikovaných podmínek užití dle poskytovatele	8
SPARQL dotaz	9
Q2: Počet datových sad, jejichž distribuce nemají specifikované podmínky užití dle poskytovatele	10
SPARQL dotaz	11
Q3: Počet záznamů datových sad nesplňujících povinné atributy dle poskytovatele	12
SPARQL dotaz pro přehled	12
SPARQL dotaz pro seznam jednotlivých datových sad	13
Q4: Formáty dat distribucí dle poskytovatele	14
SPARQL dotaz pro přehled	15
SPARQL dotaz pro seznam jednotlivých distribucí	16



Q5: Podmínky užití distribucí dle poskytovatele	16
SPARQL dotaz pro přehled	17
SPARQL dotaz pro seznam jednotlivých distribucí	17
Q5b: Počty poskytovatelů dle podmínek užití distribucí	18
SPARQL dotaz	18
Q6: Počty datových sad s danou periodicitou aktualizace dle poskytovatele	19
SPARQL dotaz pro přehled dle poskytovatelů	19
SPARQL dotaz pro celkový přehled o datových sadách	20
Q7: Počet datových sad a distribucí dle poskytovatele	20
SPARQL dotaz	21
Dostupnost zdrojů registrovaných v NKOD	21
Pipeline v LinkedPipes ETL	22
A1.1: Nedostupnost distribucí datových sad	23
SPARQL dotaz	23
A1.2: Nedostupné distribuce datových sad	25
SPARQL dotaz	26
A2.1: Nedostupnost schémat distribucí datových sad	26
SPARQL dotaz	27
A2.2: Nedostupná schémata distribucí datových sad	29
SPARQL dotaz	29
A3.1: Nedostupnost podmínek užití distribucí datových sad	30
SPARQL dotaz	31
A3.2: Nedostupné podmínky užití distribucí datových sad	33
SPARQL dotaz	33
A4.1: Nedostupnost uživatelské dokumentace datových sad	34
SPARQL dotaz	35
A4.2: Nedostupné uživatelské dokumentace datových sad	37
SPARQL dotaz	37
A5.1: Neshoda mezi formátem distribuce v NKOD a formátem indikovaným serverem	38
SPARQL dotaz pro přehled	39
SPARQL dotaz pro jednotlivé distribuce datových sad	40
Otevřenost podmínek užití	41
Příloha 1 - LinkedPipes ETL pipeline pro počítání statistik z RDF reprezentace NKOD dle DCAT-AP v1.1	42



Příloha 2 - LinkedPipes ETL pipeline pro zjišťování dostupnosti dokumentace datových sad, podmínek užití, schémat, a samotných dat distribucí datových sad

42

Pojmy

- URL - Uniform Resource Locator
- NKOD - Národní katalog otevřených dat
- OD - otevřená data
- API - Application programming interface
- XML - eXtensible Markup Language
- RDF - Resource Description Framework
- SPARQL - dotazovací jazyk nad daty v RDF
- SPARQL endpoint - přístupový bod do RDF databáze umožňující vyhodnocování dotazů
- ETL - Extract, Transform, Load
- Pipeline - datový proces v nástroji LinkedPipes ETL
- LinkedPipes ETL - nástroj pro transformaci dat pomocí opakovatelných procesů - pipeline
- DCAT-AP - Evropský standard pro záznamy v datových katalozích založený na RDF

Kvalita datových sad katalogizovaných v NKOD

Kvalita dat je klíčová pro rozvoj celého ekosystému otevřených dat v České republice i mimo ni. Pouze kvalitní otevřená data si najdou své uživatele, a tedy přinesou nějakou hodnotu. Pro podporu kvality otevřených dat slouží web <https://opendata.gov.cz>. Jsou na něm k dispozici všechny potřebné informace k poskytování kvalitních otevřených dat určené všem jednotlivým rolím účastníků se procesu přípravy otevřených dat poskytovatelem. Informace jsou navíc doplňovány špatnými i dobrými příklady z praxe. I přesto se kvalita poskytovaných otevřených dat velice různí, a je tedy nutné ji kontrolovat.

Kvalita dat je nejobecněji definována jako “připravenost k použití” a dá se měřit různými objektivními i subjektivními způsoby, automaticky i ručně. Jelikož není k dispozici lidská síla, která by každou zveřejněnou sadu ručně kontrolovala až na úroveň použitelnosti, musíme se při kontrole kvality datových sad omezit na objektivní a automatizovatelné metody založené na datech, která jsou k dispozici. Důležitou roli v datové kvalitě hraje popis dat, tedy dokumentace a metadata. Metadata ve formě záznamů o otevřených datových sadách se ukládají v Národním katalogu otevřených dat (NKOD), a jsou zároveň jediným dostupným zdrojem dat, který lze využít pro automatizovanou kontrolu kvality dat.



Obsah záznamu o datové sadě v NKOD

Popis položek záznamu o datové sadě v NKOD je lze nalézt na webu

<https://opendata.gov.cz/nastroj:ckan-api>. Jedná se o položky popisující datovou sadu, a položky popisující její distribuci.

Položky popisující datovou sadu

V následující tabulce jsou položky záznamu v NKOD popisující datovou sadu.

Název položky	Formát	Povinnost	Příklad hodnoty
Název datové sady	Řetězec	Ano	Kontroly
Popis datové sady	Řetězec	Ano	Seznam kontrol od 1.1.2012
Periodicita aktualizace	Řetězec dle ISO 8601	Ano	R/P3M
Související geografické území - typ prvku RÚIAN	Typ prvku RÚIAN	Ano	ST
Související geografické území - kód prvku RÚIAN	Řetězec	Ano	1
Klasifikace dle EuroVoc	Seznam URL oddělený mezerou	Ne	http://eurovoc.europa.eu/171
Email Kurátora dat	E-mail	Ne	verejnadata@INSTITUTE.cz
Jméno Kurátora dat	Řetězec	Ne	Jan Novák
Odkaz na dokumentaci datové sady	URL	Ne	http://www.INSTITUTE.cz/open-data/dokumentace-ds1.html
Klíčová slova	pole objektů	Ano	rozpočet, smlouvy
Dotčené časové období od	datum ve formátu xsd:date	Ne	2012-01-01
Dotčené časové období do	datum ve formátu xsd:date	Ne	2014-12-31



Položky popisující distribuci datové sady

V následující tabulce jsou položky záznamu v NKOD popisující distribuci datové sady.

Název položky	Formát	Povinnost	Příklad hodnoty
Odkaz na stažení datového zdroje	URL	Ano	http://www.INSTITUCE.cz/open-data/kontroly.trig
Formát datového zdroje	MIME typ	Ano	application/trig
Odkaz na podmínky užití	URL	Ano	https://portal.gov.cz/portal/ostatni/volny-pristup-k-ds.html
Název datového zdroje	Řetězec	Ne	Kontroly
Odkaz na strojově čitelné schéma datového zdroje	URL	Ne	http://www.INSTITUCE.cz/open-data/ontologie.trig
Formát schématu datového zdroje	MIME typ	Ne	application/trig
Dotčené časové období od	datum ve formátu xsd:date	Ne	2015-01-01
Dotčené časové období do	datum ve formátu xsd:date	Ne	2015-12-31

Cíle kontroly kvality datových sad

Cílem kontroly kvality datových sad registrovaných v NKOD je zajistit, aby zejména

1. Metadata v NKOD obsahovala alespoň všechny povinné položky záznamu,
2. URL odkazy na data, dokumentaci, podmínky použití a schémata byly aktuální a funkční,
3. formát poskytovaných dat odpovídal formátu uvedenému v metadatach,
4. metadata jednotlivých poskytovatelů byla co nejvíce konzistentní, tedy například aby stejné podmínky užití byly indikovány stejným způsobem a podobně.



Kontrola kvality záznamů o otevřených datových sadách v NKOD probíhá dvěma způsoby. Prvním způsobem je kontrola samotných metadat registrovaných v NKOD. Ta jsou pomocí XML API NKOD a nástroje LinkedPipes ETL (pipeline v [Příloze 1](#)) nejprve transformována do RDF reprezentace dle standardu DCAT-AP v1.1. Následně jsou pomocí dotazování jazykem SPARQL provedeny jednotlivé kontroly. Ty zjišťují zejména zda záznamy splňují požadavky standardu DCAT-AP v1.1 na povinné položky a dále jsou z nich počítány různé statistiky určené pro monitorování stavu NKOD, konzistence metadat a indikaci chyb k řešení.

Druhým způsobem je ověření, zda metadata reflektují skutečný stav otevřených dat u poskytovatele. To je prováděno systematickým automatizovaným ověřováním existence a možnosti stažení registrovaných dat, jejich dokumentace, schémat a podmínek užití, opět pomocí nástroje LinkedPipes ETL.

Výstupy těchto kontrol jsou publikovány na webu <https://opendata.gov.cz>, kde je vidí jednotliví poskytovatelé otevřených dat a mohou na ně reagovat. Ty, co nereagují, osloví Národní koordinátor otevřených dat s žádostí o nápravu případných problémů.

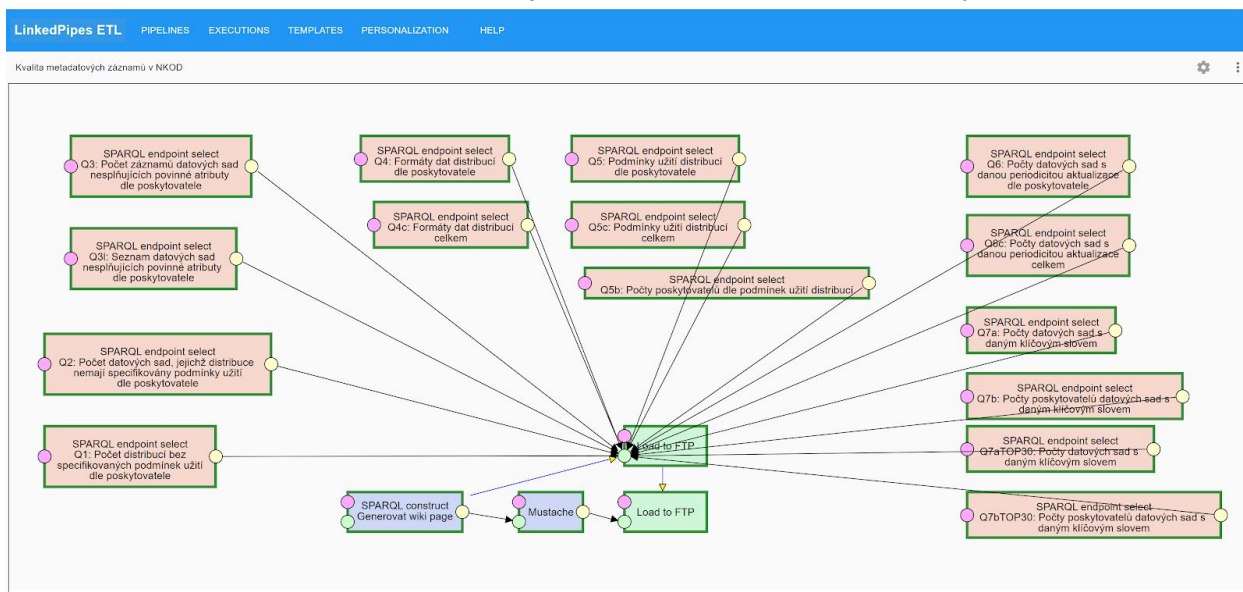


Kontrola metadat v NKOD

V této sekci popisujeme měřené statistiky a poskytujeme SPARQL dotazy pro jejich počítání.

Pipeline v nástroji LinkedPipes ETL

Na obrázku níže vidíme pipeline v nástroji LinkedPipes ETL, která statistiky počítá



Obrázek 1 - Pipeline v LinkedPipes ETL vyhodnocující SPARQL dotazy nad RDF reprezentací NKOD za účelem kontroly datové kvality

Červené komponenty znázorňují jednotlivé SPARQL dotazy kladené na SPARQL endpoint obsahující kompletní obsah NKOD transformovaný do RDF reprezentace dle standardu DCAT-AP. Po jejich spočítání dojde ve spodní části pipeline k jejich zveřejnění na webu <https://opendata.gov.cz/statistika:start> pomocí zelených komponent. V dalších sekcích popisujeme jednotlivé kontroly, jejich účel a provedení pomocí SPARQL dotazu.

V následujících částech jsou uvedeny jednotlivé přehledy včetně SPARQL dotazů, které je počítají.

Q1: Počet distribucí bez specifikovaných podmínek užití dle poskytovatele



Kód přehledu	Q1
Název přehledu	Počet distribucí bez specifikovaných podmínek užití dle poskytovatele
Výstup přehledu	Seznam poskytovatelů OD řazený sestupně dle procenta distribucí datových sad bez specifikovaných podmínek užití
Ukazatele	<ul style="list-style-type: none">• procento - podíl distribucí datových sad bez specifikovaných podmínek užití a celkového počtu distribucí datových sad poskytovatele v procentech• počet_distribucí_bez_podmínek - počet distribucí datových sad poskytovatele bez specifikovaných podmínek užití• počet_distribucí_celkem - počet distribucí datových sad poskytovatele celkem
Žádoucí hodnota	Prázdný seznam

Podmínky užití se dle standardu DCAT-AP vztahují k distribucím datových sad, jelikož každá distribuce datové sady může mít podmínky užití různé. Specifikace otevřených podmínek užití je nutnou podmínkou pro to, aby data mohla být považována za otevřená. Ideálně by se tedy neměla vyskytovat žádná distribuce datové sady bez specifikace podmínek užití. Pokud se vyskytuje, je to špatně, a počet takových výskytů monitoruje tento přehled.

SPARQL dotaz

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT ?jméno_poskytovatele ((?počet_distribucí_bez_podmínek*
100/?počet_distribucí_celkem) as ?procento) ?počet_distribucí_bez_podmínek
?počet_distribucí_celkem
WHERE
{
?poskytovatel foaf:name ?jméno_poskytovatele .
{
SELECT ?poskytovatel (count (distinct ?distribuce) as
?počet_distribucí_bez_podmínek)
WHERE {
?datová_sada a dcat:Dataset ;
```



```

dcterms:publisher ?poskytovatel ;
dcat:distribution ?distribuce .
FILTER NOT EXISTS {?distribuce dcterms:license ?license.}
}
}

{
SELECT ?poskytovatel (count (distinct ?distribuce) as
?počet_distribucí_celkem)
WHERE {
?datová_sada a dcat:Dataset ;
dcterms:publisher ?poskytovatel ;
dcat:distribution ?distribuce .
}
}

}
GROUP BY ?poskytovatel ?jméno_poskytovatele
ORDER BY DESC(?procento) ?jméno_poskytovatele

```

Q2: Počet datových sad, jejichž distribuce nemají specifikovány podmínky užití dle poskytovatele

Kód přehledu	Q2
Název přehledu	Počet datových sad, jejichž distribuce nemají specifikovány podmínky užití dle poskytovatele
Výstup přehledu	Seznam poskytovatelů OD řazený sestupně dle procenta datových sad, které mají distribuce bez specifikovaných podmínek užití
Ukazatele	<ul style="list-style-type: none"> • procento - podíl datových sad, jejichž distribuce nemají specifikovány podmínky užití a celkového počtu datových sad poskytovatele v procentech • pocet_datovych_sad_bez - počet datových sad poskytovatele, jejichž distribuce nemají specifikovány podmínky užití • pocet_datovych_sad_celkem - počet datových sad poskytovatele celkem
Žádoucí hodnota	Prázdný seznam



Podmínky užití se dle standardu DCAT-AP vztahují k distribucím datových sad, nicméně jednotkou, o které nejčastěji komunikují poskytovatelé otevřených dat je datová sada. V tomto přehledu jsou tedy počítány datové sady, které obsahují distribuci, která nemá specifikovány podmínky užití. Ideálně taková neexistuje, a pokud ano, je to špatně.

SPARQL dotaz

```
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX dcat:   <http://www.w3.org/ns/dcat#>
PREFIX foaf:   <http://xmlns.com/foaf/0.1/>

SELECT ?jméno_poskytovatele ((?pocet_datovych_sad_bez *
100/?pocet_datovych_sad_celkem) as ?procento) ?pocet_datovych_sad_bez
?pocet_datovych_sad_celkem
WHERE
{
  ?poskytovatel foaf:name ?jméno_poskytovatele .

  {
    SELECT ?poskytovatel (count (distinct ?datová_sada) as
?pocet_datovych_sad_bez)
    WHERE {
      ?datová_sada a dcat:Dataset ;
        dcterms:publisher ?poskytovatel ;
        dcat:distribution ?distribuce .
    }
    FILTER NOT EXISTS {?distribuce dcterms:license ?license.}
  }

  {
    SELECT ?poskytovatel (count (distinct ?datová_sada) as
?pocet_datovych_sad_celkem)
    WHERE {
      ?datová_sada a dcat:Dataset ;
        dcterms:publisher ?poskytovatel ;
        dcat:distribution ?distribuce .
    }
  }
}
```



```
}
}
GROUP BY ?poskytovatel ?jmeno
ORDER BY DESC(?procento) ?jmeno
```

Q3: Počet záznamů datových sad nesplňujících povinné atributy dle poskytovatele

Kód přehledu	Q3
Název přehledu	Počet záznamů datových sad nesplňujících povinné atributy dle poskytovatele
Výstup přehledu	Seznam poskytovatelů OD řazený sestupně dle počtu datových sad, které nesplňují povinné atributy dle DCAT-AP
Ukazatele	<ul style="list-style-type: none"> počet_datových_sad_nesplňujících - počet datových sad poskytovatele, které nesplňují povinné atributy dle DCAT-AP
Žádoucí hodnota	Prázdný seznam

Podmínky užití nejsou jedinou povinnou položkou metadatového záznamu datové sady. Dalšími jsou název, popis, periodicita aktualizace, související geografické území a klíčová slova popisující datovou sadu, a dále datový formát a odkaz ke stažení souboru s daty distribuce datové sady. Jejich nepřítomnost testujeme v tomto přehledu. Ideálně by takový záznam neexistoval, a pokud existuje, je to špatně. Počítáme jak přehled, tedy počet takových záznamů dle poskytovatele, který se pak objeví na webu, tak i seznam takových datových sad, aby dotčený poskytovatel věděl, které datové sady se chyba týká.

SPARQL dotaz pro přehled

```
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT DISTINCT ?jméno_poskytovatele ?počet_datových_sad_nesplňujících
WHERE
{
```



```
?poskytovatel foaf:name ?jméno_poskytovatele.  
{  
SELECT ?poskytovatel (count (distinct ?datová_sada) as  
?počet_datových_sad_nesplňujících)  
WHERE {  
?datová_sada a dcat:Dataset ;  
    dcterms:publisher ?poskytovatel .  
  
MINUS {  
?datová_sada a dcat:Dataset ;  
    dcterms:title ?title ;  
    dcterms:description ?description ;  
    dcterms:accrualPeriodicity ?periodicity ;  
    dcterms:spatial ?spatial ;  
    dcat:keyword ?keyword ;  
    dcterms:publisher ?poskytovatel ;  
    dcat:distribution ?distribution .  
  
?distribution a dcat:Distribution ;  
    dcat:accessURL ?accessUrl ;  
    dcat:downloadURL ?downloadUrl ;  
    dcat:mediaType ?mediaType ;  
    dcterms:license ?license .  
}  
}  
}  
}  
GROUP BY ?poskytovatel ?jméno_poskytovatele  
ORDER BY DESC(?počet_datových_sad_nesplňujících) ?jméno_poskytovatele
```

SPARQL dotaz pro seznam jednotlivých datových sad

```
PREFIX dcterms: <http://purl.org/dc/terms/>  
PREFIX dcat:    <http://www.w3.org/ns/dcat#>  
PREFIX foaf:    <http://xmlns.com/foaf/0.1/>  
  
SELECT DISTINCT ?jméno_poskytovatele ?datová_sada
```



```
WHERE
{
?poskytovatel foaf:name ?jméno_poskytovatele .
{
SELECT ?poskytovatel ?datová_sada
WHERE {
?datová_sada a dcat:Dataset ;
      dcterms:publisher ?poskytovatel .

MINUS {
?datová_sada a dcat:Dataset ;
      dcterms:title ?title ;
      dcterms:description ?description ;
      dcterms:accrualPeriodicity ?periodicity ;
      dcterms:spatial ?spatial ;
      dcat:keyword ?keyword ;
      dcterms:publisher ?poskytovatel ;
      dcat:distribution ?distribution .

?distribution a dcat:Distribution ;
      dcat:accessURL ?accessUrl ;
      dcat:downloadURL ?downloadUrl ;
      dcat:mediaType ?mediaType ;
      dcterms:license ?license .
}

FILTER (?datová_sada != <https://nkod.opendata.cz/zdroj/datová-sada/NKOD>)
}
}
}
GROUP BY ?poskytovatel ?jméno_poskytovatele
ORDER BY ?jméno_poskytovatele
```

Q4: Formáty dat distribucí dle poskytovatele

Kód přehledu	Q4
Název přehledu	Formáty dat distribucí dle poskytovatele



Výstup přehledu	Seznam poskytovatelů OD a formátů distribucí jejich datových sad řazený sestupně dle jejich počtu v daném formátu
Ukazatele	<ul style="list-style-type: none">• počet_distribucí - počet distribucí datových sad poskytovatele ve formátu daném sloupcem mime_type
Žádoucí hodnota	Není - jedná se o informativní přehled

Tento přehled slouží pro získání povědomí o tom, jaké datové formáty jsou v otevřených datech v NKOD zastoupeny. Také pomáhá odhalit chybně vyplněné formáty neobsažené v oficiálním seznamu MIME typů na <https://www.iana.org/assignments/media-types/>. Je počítán jak přehled podle poskytovatelů, tak seznam jednotlivých distribucí.

SPARQL dotaz pro přehled

```
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX dcat:   <http://www.w3.org/ns/dcat#>
PREFIX foaf:   <http://xmlns.com/foaf/0.1/>

SELECT ?jméno_poskytovatele (count (distinct ?distribuce) as
?počet_distribucí) ?mime_type
WHERE
{
  GRAPH ?g { ?datová_sada a dcat:Dataset ;
    dcterms:publisher ?poskytovatel ;
    dcat:distribution ?distribuce .

    ?distribuce dcat:mediaType ?mime_type.

    ?poskytovatel foaf:name ?jméno_poskytovatele .
  }
  FILTER (?g != <https://nkod.opendata.cz/zdroj/datová-sada/NKOD>)
}
GROUP BY ?poskytovatel ?jméno_poskytovatele ?mime_type
ORDER BY DESC(?počet_distribucí) ?mime_type
```



SPARQL dotaz pro seznam jednotlivých distribucí

```
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX dcat:   <http://www.w3.org/ns/dcat#>
PREFIX foaf:   <http://xmlns.com/foaf/0.1/>

SELECT ?mime_type (count (distinct ?distribuce) as ?počet_distribucí)
WHERE
{
  ?datová_sada a dcat:Dataset ;
    dcterms:publisher ?poskytovatel ;
    dcat:distribution ?distribuce .

  ?distribuce dcat:mediaType ?mime_type.
FILTER (?datová_sada != <https://nkod.opendata.cz/zdroj/datová-sada/NKOD>)
}
GROUP BY ?mime_type
ORDER BY DESC(?počet_distribucí)
```

Q5: Podmínky užití distribucí dle poskytovatele

Kód přehledu	Q5
Název přehledu	Podmínky užití distribucí dle poskytovatele
Výstup přehledu	Seznam poskytovatelů OD a podmínek užití distribucí jejich datových sad řazený sestupně dle jejich počtu s danými podmínkami užití
Ukazatele	<ul style="list-style-type: none"> počet_distribucí - počet distribucí datových sad poskytovatele s podmínkami užití danými sloupcem podmínky_užití
Žádoucí hodnota	Jedno URI pro jedny podmínky užití napříč poskytovateli

Tento přehled sleduje, jaký poskytovatel používá jaké podmínky užití pro distribuce svých datových sad. Je počítán jak přehled podle poskytovatelů, tak seznam jednotlivých distribucí. Přehled slouží jako podklad pro sledování konzistence a rozmanitosti užitých podmínek užití a jejich URI. Cílem by mělo být používání jediného URI pro stejné podmínky užití a minimalizace různých podmínek užití napříč poskytovateli otevřených dat.



SPARQL dotaz pro přehled

```
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX dcat:   <http://www.w3.org/ns/dcat#>
PREFIX foaf:   <http://xmlns.com/foaf/0.1/>

SELECT ?jméno_poskytovatele (count (distinct ?distribuce) as
?počet_distribucí) ?podmínky_užití
WHERE
{
  GRAPH ?g {
    ?datová_sada a dcat:Dataset ;
      dcterms:publisher ?poskytovatel ;
      dcat:distribution ?distribuce .

    ?distribuce dcterms:license ?podmínky_užití.

    ?poskytovatel foaf:name ?jméno_poskytovatele .
  }
  FILTER (?g != <https://nkod.opendata.cz/zdroj/datová-sada/NKOD>)
}
GROUP BY ?poskytovatel ?jméno_poskytovatele ?podmínky_užití
ORDER BY DESC(?počet_distribucí) ?podmínky_užití
```

SPARQL dotaz pro seznam jednotlivých distribucí

```
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX dcat:   <http://www.w3.org/ns/dcat#>
PREFIX foaf:   <http://xmlns.com/foaf/0.1/>

SELECT (count (distinct ?distribuce) as ?počet_distribucí) ?podmínky_užití
WHERE
{
  ?datová_sada a dcat:Dataset ;
    dcterms:publisher ?poskytovatel ;
    dcat:distribution ?distribuce .
}
```



```
?distribuce dcterms:license ?podmínky_užití.
FILTER (?datová_sada != <https://nkod.opendata.cz/zdroj/datová-sada/NKOD>)
}
GROUP BY ?podmínky_užití
ORDER BY DESC(?počet_distribucí) ?podmínky_užití
```

Q5b: Počty poskytovatelů dle podmínek užití distribucí

Kód přehledu	Q5b
Název přehledu	Počty poskytovatelů dle podmínek užití distribucí
Výstup přehledu	Seznam podmínek užití řazený sestupně dle počtu poskytovatelů OD, kteří je využívají
Ukazatele	<ul style="list-style-type: none"> počet_poskytovatelů - počet poskytovatelů OD užívající podmínky užití dané sloupcem podmínky_užití
Žádoucí hodnota	Jedno URI pro jedny podmínky užití napříč poskytovateli

Tento přehled ukazuje, kolika poskytovateli otevřených dat jsou dané podmínky užití použity. Přehled pomáhá odhalit různé specifikace (URI) stejných podmínek užití, a slouží tedy jako podklad pro zajištění jejich konzistence. Dále pomáhá určit, který poskytovatel používá vlastní podmínky užití, které je třeba zkontrolovat, zda splňují podmínku otevřenosti.

SPARQL dotaz

```
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT ?podmínky_užití (count(distinct ?poskytovatel) as
?počet_poskytovatelů)
WHERE
{
GRAPH ?g {
?datová_sada a dcat:Dataset ;
dcterms:publisher ?poskytovatel ;
dcat:distribution ?distribuce .
```



```
?distribuce dcterms:license ?podmínky_užití.
}
FILTER (?g != <https://nkod.opendata.cz/zdroj/datová-sada/NKOD>)
}
GROUP BY ?podmínky_užití
ORDER BY DESC(?počet_poskytovatelů) ?podmínky_užití
```

Q6: Počty datových sad s danou periodicitou aktualizace dle poskytovatele

Kód přehledu	Q6
Název přehledu	Počty datových sad s danou periodicitou aktualizace dle poskytovatele
Výstup přehledu	Seznam periodicit aktualizace datových sad a poskytovatelů OD, kteří poskytují OD v dané periodicitě řazený sestupně dle počtu datových sad s danou periodicitou
Ukazatele	<ul style="list-style-type: none"> počet_datových_sad - počet datových sad poskytovaných daným poskytovatelem OD ve sloupci jméno_poskytovatele v periodicitě dané sloupcem periodicita
Žádoucí hodnota	Není - jedná se o informativní přehled

Tento přehled poskytuje informaci o tom, který poskytovatel indikuje jakou periodicitu aktualizace jeho datových sad. Slouží jako základ pro detekci chybně vyplněné periodicity aktualizace a jako informativní přehled o periodicitách, s jakými chtějí poskytovatelé OD datové sady publikovat. Je zde počítán nejprve přehled pro jednotlivé poskytovatele a následně ještě přehled datových sad celkem, napříč poskytovateli.

SPARQL dotaz pro přehled dle poskytovatelů

```
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT ?periodicita ?jméno_poskytovatele (count (distinct ?datova_sada) as
?počet_datových_sad)
WHERE
```



```
{
GRAPH ?g {
?datova_sada a dcat:Dataset ;
    dcterms:publisher ?poskytovatel ;
    dcterms:accrualPeriodicity ?periodicita .

?poskytovatel foaf:name ?jméno_poskytovatele .
}
FILTER (?g != <https://nkod.opendata.cz/zdroj/datová-sada/NKOD>)
}
GROUP BY ?periodicita ?poskytovatel ?jméno_poskytovatele
ORDER BY DESC(?počet_datových_sad)
```

SPARQL dotaz pro celkový přehled o datových sadách

```
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT ?periodicita (count (distinct ?datová_sada ) as ?počet_datových_sad)
WHERE
{
?datová_sada a dcat:Dataset ;
    dcterms:publisher ?poskytovatel ;
    dcterms:accrualPeriodicity ?periodicita .
FILTER (?datová_sada != <https://nkod.opendata.cz/zdroj/datová-sada/NKOD>)
}
GROUP BY ?periodicita
ORDER BY DESC(?počet_datových_sad)
```

Q7: Počet datových sad a distribucí dle poskytovatele

Kód přehledu	Q7
Název přehledu	Počet datových sad a distribucí dle poskytovatele



Výstup přehledu	Seznam poskytovatelů OD, celkového počtu jejich datových sad a celkového počtu jejich distribucí, řazený sestupně dle počtu datových sad
Ukazatele	<ul style="list-style-type: none">● PočetDatovýchSad - počet datových sad poskytovaných daným poskytovatelem OD● PočetDistribucí - počet distribucí datových sad poskytovaných daným poskytovatelem OD
Žádoucí hodnota	Není - jedná se o informativní přehled

Počet datových sad a distribucí sám o sobě nemá žádnou výpovědní hodnotu a neměl by se používat jako indikátor ve světě otevřených dat. Nesouvisí totiž nijak s hodnotou poskytovaných dat. Měříme ho z technických důvodů pro účely monitoringu a predikce zátěže NKOD.

SPARQL dotaz

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dcterms: <http://purl.org/dc/terms/>

SELECT ?Poskytovatel (count(distinct ?dataset) as ?PočetDatovýchSad)
(count(distinct ?distribution) as ?PočetDistribucí)
WHERE {
  GRAPH [] {
    ?dataset a dcat:Dataset;
              dcterms:publisher ?publisher .
    OPTIONAL {?dataset dcat:distribution ?distribution . }
    ?publisher foaf:name ?Poskytovatel.
  }
}
GROUP BY ?Poskytovatel
ORDER BY DESC(?PočetDatovýchSad)
```

Dostupnost zdrojů registrovaných v NKOD

Krom dotazů nad metadaty v NKOD, které jsme popsali v minulé sekci, lze na jejich základě automatizovaně kontrolovat dostupnost odkazovaných zdrojů. Ty zahrnují

- samotné datové soubory

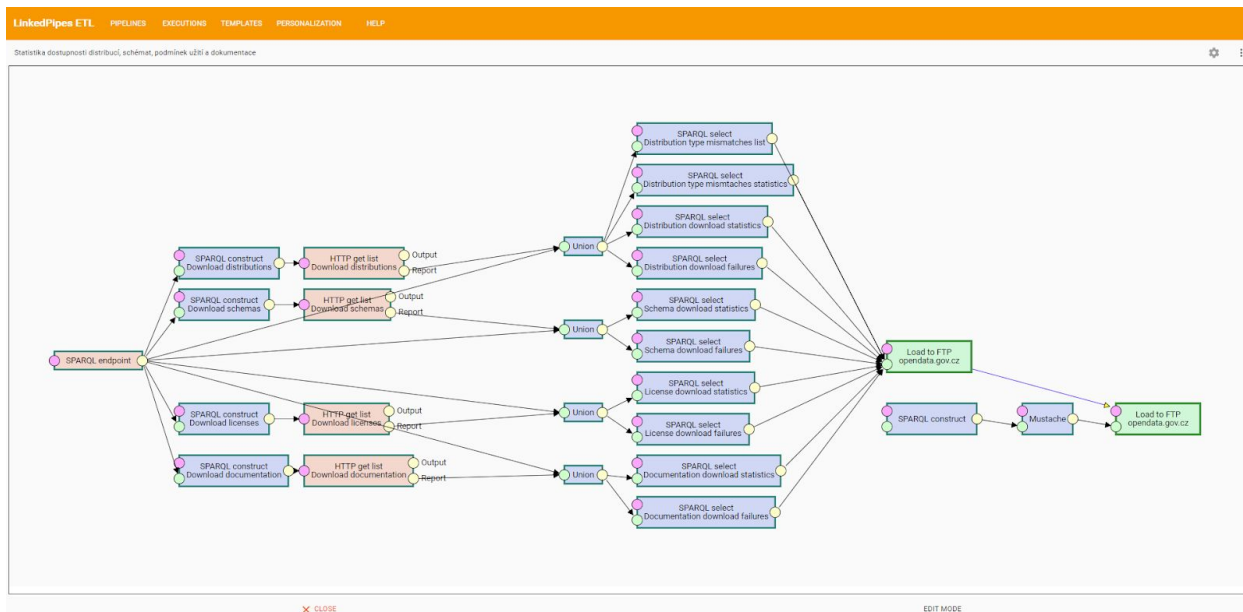


- schémata datových souborů
- dokumentace datových sad
- dokumenty obsahující podmínky užití

Tyto zdroje jsou v NKOD registrovány pomocí svého URL. Kontrola tedy probíhá přístupem na tuto URL pomocí metody HTTP HEAD (získání hlaviček) a HTTP GET (získání hlaviček i obsahu). Zejména je kontrolována serverem poskytovatele zasláná hlavička Content-Type určující formát dat. Nad ní pak probíhá kontrola zjišťující, zda se shoduje se záznamem v NKOD. Kontrola opět probíhá pomocí pipeline - procesu v nástroji LinkedPipes ETL.

Pipeline v LinkedPipes ETL

Pipeline se skládá z několika základních kroků. V levé části pipeline (v [Příloze 2](#)) je komponenta získávající potřebná metadata ze SPARQL endpointu obsahující RDF reprezentaci NKOD dle standardu DCAT-AP. Na základě takto získaných metadat je vygenerována konfigurace (modré komponenty) pro komponenty přistupující na daná URL (červené komponenty - <https://etl.linkedpipes.com/components/e-httpgetfiles>). Tyto komponenty pak generují zprávu obsahující výsledek přístupu na každé URL, zejména stavový kód HTTP určující, zda byl zdroj nalezen, nebo zda došlo k nějakému problému. Dále zpráva obsahuje právě HTTP hlavičky zasláné serverem. V případě nastalého problému dále zpráva obsahuje jeho popis. Tyto zprávy jsou dále zpracovány do požadovaných přehledů a seznamů (modré komponenty), které jsou následně zveřejněny na webu <https://opendata.gov.cz/statistika:start> (zelené komponenty).



Obrázek 2 - Pipeline v LinkedPipes ETL vyhodnocující dostupnost zdrojů registrovaných v NKOD

V následujících sekcích popisujeme vzniklé přehledy dostupnosti.



A1.1: Nedostupnost distribucí datových sad

Kód přehledu	A1.1
Název přehledu	Nedostupnost distribucí datových sad
Výstup přehledu	Seznam poskytovatelů OD, kteří mají registrovány nedostupné distribuce datových sad, jejich počet, podíl vůči celkovému počtu distribucí datových sad a podíl vůči celkovému počtu distribucí datových sad s odkazem ke stažení. Řazeno sestupně dle podílu vůči celkovému počtu distribucí datových sad.
Ukazatele	<ul style="list-style-type: none">• procento_nedostupných_ze_všech - Procento nedostupných distribucí ze všech distribucí• procento_nedostupných_z_ke_stažení - Procento nedostupných distribucí z distribucí, které mají odkaz ke stažení• distribucí - Počet distribucí celkem• ke_stažení - Počet distribucí s odkazem ke stažení celkem• nedostupných - Počet nedostupných distribucí
Žádoucí hodnota	Prázdný seznam - žádné nedostupné distribuce datových sad

V tomto přehledu jsou uvedeni ti poskytovatelé otevřených dat, jejichž distribuce datových sad jsou z jakéhokoliv důvodu v době kontroly nedostupné. Je třeba podotknout, že z důvodu chybné registrace mohou existovat i distribuce datových sad, které nemají odkaz ke stažení.

SPARQL dotaz

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX report: <https://vocabulary.etl.linkedpipes.com/report/>
PREFIX httpList: <http://plugins.linkedpipes.com/ontology/e-httpGetFiles#>

SELECT (MIN(?pName) AS ?jméno_poskytovatele)
(xsd:integer((ceil(?nedostupných*100/?distribucí)))) as
?procento_nedostupných_ze_všech)
```



```
(xsd:integer((ceil(?nedostupných*100/?ke_stažení))) as
?procento_nedostupných_z_ke_stažení) ?distribucí ?ke_stažení ?nedostupných
WHERE
{
?poskytovatel foaf:name ?pName .
FILTER(LANG(?pName) = "cs")

{
SELECT ?poskytovatel (COUNT (DISTINCT ?distribuce) as ?nedostupných)
WHERE
{
?datová_sada a dcat:Dataset;
  dcat:distribution ?distribuce ;
  dcterms:publisher ?poskytovatel .
?distribuce dcat:downloadURL ?downloadURL .
?report report:task ?downloadURL ;
  report:status report:Failed ;
  report:duration ?duration .

OPTIONAL {
?report report:exception ?exception .
?exception report:message ?message .
}
}
GROUP BY ?poskytovatel
}

{
SELECT ?poskytovatel (COUNT (DISTINCT ?distribuce) as ?ke_stažení)
WHERE
{
?datová_sada a dcat:Dataset;
  dcat:distribution ?distribuce ;
  dcterms:publisher ?poskytovatel .
?distribuce dcat:downloadURL ?downloadURL .
?report report:task ?downloadURL ;
  report:status ?status ;
  report:duration ?duration .
```




```
OPTIONAL {
?report report:exception ?exception .
?exception report:message ?message .
}
}
GROUP BY ?poskytovatel
}

{
SELECT ?poskytovatel (COUNT (DISTINCT ?distribuce) as ?distribucí)
WHERE
{
?datová_sada a dcat:Dataset;
    dcat:distribution ?distribuce ;
    dcterms:publisher ?poskytovatel .
}
GROUP BY ?poskytovatel
}
}
GROUP BY ?poskytovatel ?distribucí ?nedostupných ?ke_stažení
ORDER BY DESC(?procento_nedostupných_ze_vsech)
```

A1.2: Nedostupné distribuce datových sad

Kód přehledu	A1.2
Název přehledu	Nedostupné distribuce datových sad
Výstup přehledu	Seznam distribucí datových sad, u kterých byl problém s dostupností, jejich URL a popis problému.
Ukazatele	<ul style="list-style-type: none">chyba - Chyba kvůli které je distribuce nedostupnásoubor - URL nedostupné distribuce
Žádoucí hodnota	Prázdný seznam - žádné nedostupné distribuce datových sad

V tomto přehledu je uveden seznam distribucí datových sad, u kterých byl v době kontroly jakýkoliv problém s dostupností.



SPARQL dotaz

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

PREFIX report: <https://vocabulary.etl.linkedpipes.com/report/>
PREFIX httpList: <http://plugins.linkedpipes.com/ontology/e-httpGetFiles#>

SELECT (MIN(?pName) AS ?jméno_poskytovatele) ?chyba ?soubor
WHERE
{
  ?s a dcat:Dataset;
      dcat:distribution ?distribution ;
      dcterms:publisher ?publisher .
  ?distribution dcat:downloadURL ?soubor.
  ?report report:task ?soubor;
      report:status report:Failed ;
      report:duration ?duration ;
      report:exception ?exception .

  ?exception report:message ?chyba.

  ?publisher foaf:name ?pName.
  FILTER(LANG(?pName) = "cs")
}
GROUP BY ?publisher ?soubor ?chyba
ORDER BY ?jméno_poskytovatele
```

A2.1: Nedostupnost schémat distribucí datových sad

Kód přehledu	A2.1
Název přehledu	Nedostupnost schémat distribucí datových sad
Výstup přehledu	Seznam poskytovatelů OD, kteří mají registrovány distribuce datových sad s nedostupným schématem, jejich počet, podíl vůči celkovému počtu distribucí datových sad a podíl vůči celkovému počtu distribucí datových sad s registrovaným schématem. Řazeno sestupně dle podílu



	distribucí s nedostupným schématem vůči celkovému počtu distribucí datových sad.
Ukazatele	<ul style="list-style-type: none">• procento_nedostupných_ze_všech - Procento distribucí s nedostupným schématem ze všech distribucí• procento_nedostupných_z_se_schématem - Procento distribucí s nedostupným schématem z distribucí, které mají schéma• distribucí - Počet distribucí celkem• se_schématem - Počet distribucí se schématem celkem• nedostupných - Počet nedostupných distribucí
Žádoucí hodnota	Prázdný seznam - žádné distribuce datových sad s nedostupným schématem

V tomto přehledu jsou uvedeni ti poskytovatelé otevřených dat, jejichž schémata distribucí datových sad jsou z jakéhokoliv důvodu v době kontroly nedostupná. Je třeba podotknout, že ne každá distribuce má datové schéma, i když by tomu tak mělo být.

SPARQL dotaz

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX report: <https://vocabulary.etl.linkedpipes.com/report/>
PREFIX httpList: <http://plugins.linkedpipes.com/ontology/e-httpGetFiles#>

SELECT (MIN(?pName) AS ?jméno_poskytovatele)
(xsd:integer((ceil(?nedostupných*100/?distribucí)))) as
?procento_nedostupných_ze_všech)
(xsd:integer((ceil(?nedostupných*100/?se_schématem)))) as
?procento_nedostupných_z_se_schématem) ?distribucí ?se_schématem
?nedostupných
WHERE
{
?poskytovatel foaf:name ?pName .
FILTER(LANG(?pName) = "cs")
```



```
{
SELECT ?poskytovatel (COUNT (DISTINCT ?distribuce) as ?nedostupných)
WHERE
{
?datová_sada a dcat:Dataset;
  dcat:distribution ?distribuce ;
  dcterms:publisher ?poskytovatel .
?distribuce dcterms:conformsTo ?downloadURL .
?report report:task ?downloadURL ;
  report:status report:Failed ;
  report:duration ?duration .

OPTIONAL {
?report report:exception ?exception .
?exception report:message ?message .
}
}
GROUP BY ?poskytovatel
}

{
SELECT ?poskytovatel (COUNT (DISTINCT ?distribuce) as ?se_schématem)
WHERE
{
?datová_sada a dcat:Dataset;
  dcat:distribution ?distribuce ;
  dcterms:publisher ?poskytovatel .
?distribuce dcterms:conformsTo ?downloadURL .
?report report:task ?downloadURL ;
  report:status ?status ;
  report:duration ?duration .

OPTIONAL {
?report report:exception ?exception .
?exception report:message ?message .
}
}
GROUP BY ?poskytovatel
}
```



```
}  
  
{  
SELECT ?poskytovatel (COUNT (DISTINCT ?distribuce) as ?distribucí)  
WHERE  
{  
?datová_sada a dcat:Dataset;  
  dcat:distribution ?distribuce ;  
  dcterms:publisher ?poskytovatel .  
}  
GROUP BY ?poskytovatel  
}  
}  
  
GROUP BY ?poskytovatel ?distribucí ?nedostupných ?se_schématem  
ORDER BY DESC(?procento_nedostupných_ze_všech)
```

A2.2: Nedostupná schémata distribucí datových sad

Kód přehledu	A2.2
Název přehledu	Nedostupná schémata distribucí datových sad
Výstup přehledu	Seznam schémat distribucí datových sad, u kterých byl problém s dostupností, jejich URL a popis problému.
Ukazatele	<ul style="list-style-type: none">chyba - Chyba kvůli které je schéma nedostupnéschéma - URL nedostupného schématu
Žádoucí hodnota	Prázdný seznam - žádné distribuce datových sad s nedostupným schématem

V tomto přehledu je uveden seznam schémat distribucí datových sad, u kterých byl v době kontroly jakýkoliv problém s dostupností.

SPARQL dotaz

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>  
PREFIX dcterms: <http://purl.org/dc/terms/>  
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
```



```
PREFIX report: <https://vocabulary.etl.linkedpipes.com/report/>
PREFIX httpList: <http://plugins.linkedpipes.com/ontology/e-httpGetFiles#>

SELECT (MIN(?pName) AS ?jméno_poskytovatele) ?chyba ?schéma
WHERE
{
  ?s a dcat:Dataset;
    dcat:distribution ?distribution ;
    dcterms:publisher ?publisher .
  ?distribution dcterms:conformsTo ?schéma.
  ?report report:task ?schéma;
    report:status report:Failed ;
    report:duration ?duration ;
    report:exception ?exception .

  ?exception report:message ?chyba.

  ?publisher foaf:name ?pName.
  FILTER(LANG(?pName) = "cs")
}
GROUP BY ?publisher ?schéma ?chyba
ORDER BY ?jméno_poskytovatele
```

A3.1: Nedostupnost podmínek užití distribucí datových sad

Kód přehledu	A3.1
Název přehledu	Nedostupnost podmínek užití distribucí datových sad
Výstup přehledu	Seznam poskytovatelů OD, kteří mají registrovány distribuce datových sad s nedostupnými podmínkami užití, jejich počet, podíl vůči celkovému počtu distribucí datových sad a podíl vůči celkovému počtu distribucí datových sad s registrovanými podmínkami užití. Řazeno sestupně dle podílu distribucí s nedostupnými podmínkami užití vůči celkovému počtu distribucí datových sad.
Ukazatele	<ul style="list-style-type: none">procento_nedostupných_ze_všech - Procento distribucí s nedostupnými podmínkami užití ze všech distribucí



	<ul style="list-style-type: none">• procento_nedostupných_z_s_podmínkami_užití - Procento distribucí s nedostupnými podmínkami užití z distribucí, které mají podmínky užití• distribucí - Počet distribucí celkem• s_podmínkami_užití - Počet distribucí s podmínkami užití celkem• nedostupných - Počet nedostupných distribucí
Žádoucí hodnota	Prázdný seznam - žádné distribuce datových sad s nedostupnými podmínkami užití

V tomto přehledu jsou uvedeni ti poskytovatelé otevřených dat, jejichž podmínky užití distribucí datových sad jsou z jakéhokoliv důvodu v době kontroly nedostupné. Je třeba podotknout, že z důvodu neúplných záznamů nemusí mít distribuce vyplněny podmínky užití, i když by tomu tak mělo být.

SPARQL dotaz

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX report: <https://vocabulary.etl.linkedpipes.com/report/>
PREFIX httpList: <http://plugins.linkedpipes.com/ontology/e-httpGetFiles#>

SELECT (MIN(?pName) AS ?jméno_poskytovatele)
(xsd:integer((ceil(?nedostupných*100/?distribucí)))) as
?procento_nedostupných_ze_všech)
(xsd:integer((ceil(?nedostupných*100/?s_podmínkami_užití)))) as
?procento_nedostupných_z_s_podmínkami_užití) ?distribucí ?s_podmínkami_užití
?nedostupných
WHERE
{
?poskytovatel foaf:name ?pName .
FILTER(LANG(?pName) = "cs")

{
SELECT ?poskytovatel (COUNT (DISTINCT ?distribuce) as ?nedostupných)
```



```
WHERE
{
?datová_sada a dcat:Dataset;
    dcat:distribution ?distribuce ;
    dcterms:publisher ?poskytovatel .
?distribuce dcterms:license ?downloadURL .
OPTIONAL {
?report report:task ?downloadURL ;
    report:status report:Failed ;
    report:duration ?duration .

OPTIONAL {
?report report:exception ?exception .
?exception report:message ?message .
}
}
FILTER(BOUND(?report))
}
GROUP BY ?poskytovatel
}

{
SELECT ?poskytovatel (COUNT (DISTINCT ?distribuce) as ?s_podmínkami_užití)
WHERE
{
?datová_sada a dcat:Dataset;
    dcat:distribution ?distribuce ;
    dcterms:publisher ?poskytovatel .
?distribuce dcterms:license ?downloadURL .
OPTIONAL {
?report report:task ?downloadURL ;
    report:status report:Failed ;
    report:duration ?duration .

OPTIONAL {
?report report:exception ?exception .
?exception report:message ?message .
}
}
}
```




```
FILTER(BOUND(?report))
}
GROUP BY ?poskytovatel
}
{
SELECT ?poskytovatel (COUNT (DISTINCT ?distribuce) as ?distribucí)
WHERE
{
?datová_sada a dcat:Dataset;
    dcat:distribution ?distribuce ;
    dcterms:publisher ?poskytovatel .
}
GROUP BY ?poskytovatel
}
}
GROUP BY ?poskytovatel ?distribucí ?nedostupných ?s_podmínkami_užití
ORDER BY DESC(?procento_nedostupných_ze_všech)
```

A3.2: Nedostupné podmínky užití distribucí datových sad

Kód přehledu	A3.2
Název přehledu	Nedostupné podmínky užití distribucí datových sad
Výstup přehledu	Seznam podmínek užití distribucí datových sad, u kterých byl problém s dostupností, jejich URL a popis problému.
Ukazatele	<ul style="list-style-type: none">chyba - Chyba kvůli které jsou podmínky užití nedostupnépodmínky_užití - URL nedostupných podmínek užití
Žádoucí hodnota	Prázdný seznam - žádné distribuce datových sad s nedostupnými podmínkami užití

V tomto přehledu je uveden seznam podmínek užití distribucí datových sad, u kterých byl v době kontroly jakýkoliv problém s dostupností.

SPARQL dotaz



```
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

PREFIX report: <https://vocabulary.etl.linkedpipes.com/report/>
PREFIX httpList: <http://plugins.linkedpipes.com/ontology/e-httpGetFiles#>

SELECT (MIN(?pName) AS ?jméno_poskytovatele) ?chyba ?podmínky_užití
WHERE
{
?s a dcat:Dataset;
    dcat:distribution ?distribution ;
    dcterms:publisher ?publisher .
?distribution dcterms:license ?podmínky_užití.
OPTIONAL {
?report report:task ?podmínky_užití;
    report:status report:Failed ;
    report:duration ?duration ;
    report:exception ?exception .

?exception report:message ?chyba.
}

?publisher foaf:name ?pName.
FILTER(LANG(?pName) = "cs")
FILTER(BOUND(?report))
}
GROUP BY ?publisher ?podmínky_užití ?chyba
ORDER BY ?jméno_poskytovatele
```

A4.1: Nedostupnost uživatelské dokumentace datových sad

Kód přehledu	A4.1
Název přehledu	Nedostupnost uživatelské dokumentace datových sad
Výstup přehledu	Seznam poskytovatelů OD, kteří mají registrovány datové sady s nedostupnou dokumentací, jejich počet, podíl vůči celkovému počtu datových sad a podíl vůči celkovému počtu datových sad s registrovanou dokumentací. Řazeno sestupně dle podílu datových sad



	s nedostupnou dokumentací vůči celkovému počtu datových sad.
Ukazatele	<ul style="list-style-type: none">• procento_nedostupných_ze_všech - Procento datových sad s nedostupnou dokumentací ze všech datových sad• procento_nedostupných_z_s_dokumentací - Procento datových sad s nedostupnou dokumentací z datových sad, které dokumentaci mají• distribucí - Počet datových sad celkem• s_dokumentací - Počet datových sad s dokumentací celkem• nedostupných - Počet datových sad s nedostupnou dokumentací
Žádoucí hodnota	Prázdný seznam - žádné datové sady s nedostupnou dokumentací

V tomto přehledu jsou uvedeni ti poskytovatelé otevřených dat, jejichž dokumentace datových sad jsou z jakéhokoliv důvodu v době kontroly nedostupné. Je třeba podotknout, dokumentace datové sady není povinná, i když je silně doporučena.

SPARQL dotaz

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX report: <https://vocabulary.etl.linkedpipes.com/report/>
PREFIX httpList: <http://plugins.linkedpipes.com/ontology/e-httpGetFiles#>

SELECT (MIN(?pName) AS ?jméno_poskytovatele)
(xsd:integer((ceil(?nedostupných*100/?datových_sad)))) as
?procento_nedostupných_ze_všech)
(xsd:integer((ceil(?nedostupných*100/?s_dokumentací)))) as
?procento_nedostupných_z_s_dokumentací) ?datových_sad ?s_dokumentací
?nedostupných
WHERE
{
?poskytovatel foaf:name ?pName .
FILTER(LANG(?pName) = "cs")
```



```
{
SELECT ?poskytovatel (COUNT (DISTINCT ?datová_sada) as ?nedostupných)
WHERE
{
?datová_sada a dcat:Dataset;
  dcat:distribution ?distribuce ;
  dcterms:publisher ?poskytovatel ;
  foaf:page ?downloadURL .
?report report:task ?downloadURL ;
  report:status report:Failed ;
  report:duration ?duration .

OPTIONAL {
?report report:exception ?exception .
?exception report:message ?message .
}
}
GROUP BY ?poskytovatel
}

{
SELECT ?poskytovatel (COUNT (DISTINCT ?datová_sada) as ?s_dokumentací)
WHERE
{
?datová_sada a dcat:Dataset;
  dcat:distribution ?distribuce ;
  dcterms:publisher ?poskytovatel ;
  foaf:page ?downloadURL .
?report report:task ?downloadURL ;
  report:status ?status ;
  report:duration ?duration .

OPTIONAL {
?report report:exception ?exception .
?exception report:message ?message .
}
}
GROUP BY ?poskytovatel
}
```



```
{  
SELECT ?poskytovatel (COUNT (DISTINCT ?datová_sada) as ?datových_sad)  
WHERE  
{  
{  
?datová_sada a dcat:Dataset;  
  dcterms:publisher ?poskytovatel .  
}  
}  
GROUP BY ?poskytovatel  
}  
}  
GROUP BY ?poskytovatel ?datových_sad ?nedostupných ?s_dokumentací  
ORDER BY DESC(?procento_nedostupných_ze_všech)
```

A4.2: Nedostupné uživatelské dokumentace datových sad

Kód přehledu	A4.2
Název přehledu	Nedostupné uživatelské dokumentace datových sad
Výstup přehledu	Seznam dokumentací datových sad, u kterých byl problém s dostupností, jejich URL a popis problému.
Ukazatele	<ul style="list-style-type: none">● chyba - Chyba kvůli které je dokumentace nedostupná● dokumentace - URL nedostupné dokumentace
Žádoucí hodnota	Prázdný seznam - žádné datové sady s nedostupnou dokumentací

V tomto přehledu je uveden seznam dokumentací datových sad, u kterých byl v době kontroly jakýkoliv problém s dostupností.

SPARQL dotaz

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>  
PREFIX dcterms: <http://purl.org/dc/terms/>  
PREFIX foaf: <http://xmlns.com/foaf/0.1/>  
  
PREFIX report: <https://vocabulary.etl.linkedpipes.com/report/>
```



```
SELECT (MIN(?pName) AS ?jméno_poskytovatele) ?chyba ?dokumentace
WHERE
{
  ?s a dcat:Dataset;
    dcat:distribution ?distribution ;
    dcterms:publisher ?publisher ;
    foaf:page ?dokumentace .
  ?report report:task ?dokumentace;
    report:status report:Failed ;
    report:duration ?duration ;
    report:exception ?exception .

  ?exception report:message ?chyba.

  ?publisher foaf:name ?pName.
  FILTER(LANG(?pName) = "cs")
}
GROUP BY ?publisher ?dokumentace ?chyba
ORDER BY ?jméno_poskytovatele
```

A5.1: Neshoda mezi formátem distribuce v NKOD a formátem indikovaným serverem

Kód přehledu	A5.1
Název přehledu	Neshoda mezi formátem distribuce v NKOD a formátem indikovaným serverem
Výstup přehledu	Seznam poskytovatelů OD, jejichž servery při přístupu k souborům ke stažení indikují jiný MIME typ než je registrován v NKOD
Ukazatele	<ul style="list-style-type: none">• typ_v_nkod - MIME typ uvedený v NKOD• typ_na_serveru - MIME typ uvedený HTTP serverem• distribucí - Počet distribucí s tímto typem neshody
Žádoucí hodnota	Prázdný seznam - žádné distribuce s různým MIME typem v NKOD a indikovaným serverem



V tomto přehledu jsou uvedeni poskytovatelé, jejichž servery při přístupu na URL souboru ke stažení distribuce datové sady uvádí jiný datový typ v Content-Type HTTP hlavičce, než-li je uveden jako datový formát distribuce v záznamu datové sady v NKOD. V HTTP hlavičce může být kromě datového formátu uvedeno ještě kódování textových formátů, které se do NKOD neuvádí. To je pro porovnání ignorováno.

SPARQL dotaz pro přehled

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX report: <https://vocabulary.etl.linkedpipes.com/report/>
PREFIX request: <http://plugins.linkedpipes.com/ontology/e-httpGetFiles#>

SELECT ?jméno_poskytovatele ?typ_v_nkod ?typ_na_serveru ?distribucí
WHERE
{
  ?poskytovatel foaf:name ?jméno_poskytovatele .
  FILTER(LANG(?jméno_poskytovatele) = "cs")
  FILTER(?stejný = false)

  {
    SELECT ?poskytovatel ?typ_v_nkod ?typ_na_serveru (MAX(?equals) as ?stejný)
    (COUNT (DISTINCT ?distribuce) as ?distribucí)
    WHERE
    {
      ?datová_sada a dcat:Dataset;
        dcat:distribution ?distribuce ;
        dcterms:publisher ?poskytovatel .
      ?distribuce dcat:downloadURL ?downloadURL ;
        dcat:mediaType ?mediaType .
      ?report report:task ?downloadURL ;
        report:status report:Success ;
        report:duration ?duration ;
        request:response ?response .
      ?response request:hasHeaderObject ?headerObject .
      ?headerObject request:name "Content-Type" ;
        request:value ?contentType .
    }
  }
}
```



```

BIND(STRFTER(STR(?mediaType),"http://www.iana.org/assignments/media-types/
") as ?typ_v_nkod)
BIND(REPLACE(?contentType,"([^;]+);?.*", "$1") AS ?typ_na_serveru)
BIND((?typ_v_nkod = ?typ_na_serveru ) as ?equals)
}
GROUP BY ?poskytovatel ?typ_v_nkod ?typ_na_serveru
}
}
ORDER BY ?jméno_poskytovatele

```

SPARQL dotaz pro jednotlivé distribuce datových sad

```

PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX report: <https://vocabulary.etl.linkedpipes.com/report/>
PREFIX request: <http://plugins.linkedpipes.com/ontology/e-httpGetFiles#>

SELECT ?jméno_poskytovatele ?downloadURL ?typ_v_nkod ?typ_na_serveru
WHERE
{
?poskytovatel foaf:name ?jméno_poskytovatele .
?datová_sada a dcat:Dataset;
    dcat:distribution ?distribuce ;
    dcterms:publisher ?poskytovatel .
?distribuce dcat:downloadURL ?downloadURL ;
    dcat:mediaType ?mediaType .
?report report:task ?downloadURL ;
    report:status report:Success ;
    report:duration ?duration ;
    request:response ?response .
?response request:hasHeaderObject ?headerObject .
?headerObject request:name "Content-Type" ;
    request:value ?contentType .

```




```

BIND(STRAFTER(STR(?mediaType), "http://www.iana.org/assignments/media-types/
") as ?typ_v_nkod)
BIND(REPLACE(?contentType, "([^;]+);?.*", "$1") AS ?typ_na_serveru)
BIND((?typ_v_nkod = ?typ_na_serveru ) as ?equals)

FILTER(LANG(?jméno_poskytovatele) = "cs")
FILTER(?equals = false)
}
ORDER BY ?jméno_poskytovatele

```

Otevřenost podmínek užití

Návod na to, jak stanovit podmínky užití datových sad je uveden na webu <https://opendata.gov.cz/cinnost:stanoveni-podminek-uziti>. Jelikož si každý poskytovatel otevřených dat může podmínky užití určit samostatně a zároveň ne každý poskytovatel postupuje dle uvedeného návodu, některé podmínky užití datových sad nesplňují podmínku otevřenosti, jelikož specifikuje takové podmínky, které jsou s otevřeností v rozporu. To představuje zásadní překážku v použití takových dat. Kontrolu uvedených podmínek užití, pokud tedy nejsou použity některé standardní dle návodu, může udělat pouze právník, a to ručně. Vzhledem k důležitosti otevřenosti podmínek užití je tato pravidelná kontrola výrazně doporučována. Data poskytovaná pod neotevřenými podmínkami užití v NKOD totiž způsobují zmatek a právní nejistotu pro zájemce otevřených dat.



Příloha 1 - LinkedPipes ETL pipeline pro počítání statistik z RDF reprezentace NKOD dle DCAT-AP v1.1

https://opendata.gov.cz/_media/statistika:kvalita-nkod.jsonld

Příloha 2 - LinkedPipes ETL pipeline pro zjišťování dostupnosti dokumentace datových sad, podmínek užití, schémat, a samotných dat distribucí datových sad

https://opendata.gov.cz/_media/statistika:dostupnost-nkod.jsonld