



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost



MINISTERSTVO VNITRA
ČESKÉ REPUBLIKY

C2V2

Návrh zajištění kompatibility NKOD s novými verzemi EU standardů pro katalogizaci otevřených dat

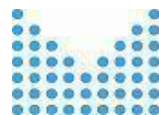
Vytvořeno v rámci projektu

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



Analýza potřebných změn oproti stávající verzi NKOD a návrh potřebných úprav

Stávající Národní katalog otevřených dat (NKOD) implementuje, co do položek metadat jednotlivých datových sad a distribucí, doporučení Evropské komise DCAT-AP 1.2.¹ založené na webovém standardu Data Catalog Vocabulary (DCAT)² vydaném konsorciem W3C. Standard DCAT byl aktualizován na verzi 2³, a v návaznosti na tuto změnu bylo aktualizováno i doporučení Evropské komise DCAT-AP na verzi 2.0.1⁴. Na základě vývoje standardů byla pro účely rozšíření metadatového záznamu datové sady vydána nová verze Otevřené formální normy (OFN) Rozhraní katalogů otevřených dat: DCAT-AP-CZ⁵ popisující všechny položky rozšířeného metadatového záznamu datové sady. Zároveň s vyvíjejícími se standardy byla sbírána i zpětná vazba od uživatelů NKOD, která vyústila v další požadavky na rozšíření NKOD. V této sekci uvádíme přehled potřebných změn v NKOD, které pochází z těchto dvou zdrojů.

Pro každý požadavek na změny je uveden i návrh, jak změny implementovat v NKOD. Typicky je třeba upravit samotné pipeliney v nástroji LinkedPipes ETL, které zajišťují harvestování lokálních katalogů a registračních záznamů zaslaných do NKOD. Dále je potřeba upravit zadávací formulář pro metadata datových sad a také zobrazení datových sad, distribucí a datových služeb v NKOD.

Umožnění registrace datových služeb

Analýza

V DCAT 2 a DCAT-AP 2.0.1 je nově možné kromě souborů ke stažení registrovat jako distribuci datové sady i tzv. datovou službu, či API. Příkladem takové datové služby může být SPARQL Endpoint, kterému je možné zaslat dotaz v jazyce SPARQL k vyhodnocení. Jedná se o principiálně odlišný způsob přístupu k datům a proto potřebuje separátní podporu v datovém katalogu.

Konkrétně je tedy třeba přidat podporu pro využití třídy Datová služba⁶ včetně jejích položek, a vazby "přístupová služba"⁷, která ji napojí na distribuci datové sady tak, jak jsou popsány v OFN.

¹ <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe/release/121>

² <https://www.w3.org/TR/2014/REC-vocab-dcat-20140116/>

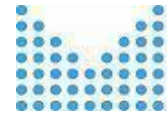
³ <https://www.w3.org/TR/vocab-dcat-2/>

⁴ <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/news/dcat-ap-release-201>

⁵ <https://ofn.gov.cz/rozhraní-katalogů-otevřených-dat/2021-01-11/>

⁶ <https://ofn.gov.cz/rozhraní-katalogů-otevřených-dat/2021-01-11/#třída-datová-slужba>

⁷ <https://ofn.gov.cz/rozhraní-katalogů-otevřených-dat/2021-01-11/#položky-distribuce-přístupová-slужba>



Návrh úprav

O příslušné položky bude rozšířena pipeline v LinkedPipes ETL zajišťující harverstaci lokálních katalogů a přípravu dat NKOD.

Bude rozšířen formulář pro zadávání metadat datových sad následujícím způsobem:

Přístup k distribuci

Typ distribuce			
Datová služba		?	
URL endpointu		?	
Odkaz na popis endpointu.	?	Odkaz na specifikaci služby (volitelné)	?
Název distribuce česky	?	Název distribuce anglicky (volitelné)	?

Dále bude implementováno zobrazování metadat datové služby na místě distribuce datové sady včetně položek dle OFN takto:

SPARQL Endpoint

Podmínky užití distribuce	Datová služba
Neobsahuje Autorské dílo ✓	Popis endpointu ✓ HTTP
Neobsahuje Originální databáze ✓	Endpoint ✓ HTTP
Není chráněna Zvláštní právo pořizovatele databáze ✓	Specifikace ✓ HTTP
Neobsahuje Osobní údaje	



Rozšíření metadatového záznamu o položky nově definované v DCAT a DCAT-AP

Analýza

V nových verzích standardů se objevily položky, které lze s výhodou použít i v českém prostředí. Jedná se o časové rozlišení datové sady⁸ a prostorové rozlišení datové sady v metrech⁹, které umožňují lépe popsat datové sady obsahující časové řady či ortofotomapy. Dále se jedná o položky distribuce datové sady lépe popisující formát souboru ke stažení, konkrétně Media type kompresního formátu¹⁰ a Media type balíčkovacího formátu¹¹.

Návrh úprav

O příslušné položky bude rozšířena pipeline v LinkedPipes ETL zajišťující harverstaci lokálních katalogů a přípravu dat NKOD.

Bude rozšířen formulář pro zadávání metadat datových sad následujícím způsobem:

Časové rozlišení (volitelné) ? Prostorové rozlišení v metrech (volitelné) ?

Dále bude rozšířen v oblasti zadávání metadat distribucí datových sad ve formě souboru ke stažení následujícím způsobem:

Přístup k distribuci

Typ distribuce
➔ Soubor ke stažení ?

Vyberte zda se jedná o stažitelný soubor nebo webovou službu

➕ Odkaz na soubor ke stažení ?

📄 Formát souboru ke stažení ? 📄 Media type souboru ke stažení ?

📄 Media type komprese (volitelné) ? 📄 Media type balíčku (volitelné) ?

↔ Odkaz na schéma distribuce (volitelné) ?

🇨🇵 Název distribuce česky (volitelné) ? 🇬🇧 Název distribuce anglicky (volitelné) ?

Zobrazení v NKOD na úrovni datové sady bude rozšířeno následujícím způsobem:

⁸ <https://ofn.gov.cz/rozhraní-katalogů-otevřených-dat/2021-01-11/#položky-datová-sada-časové-rozlišení>

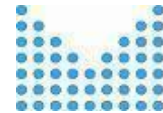
⁹ <https://ofn.gov.cz/rozhraní-katalogů-otevřených-dat/2021-01-11/#položky-datová-sada-prostorové-rozlišení-v-metrech>

¹⁰ <https://ofn.gov.cz/rozhraní-katalogů-otevřených-dat/2021-01-11/#položky-distribuce-kompresní-formát>

¹¹ <https://ofn.gov.cz/rozhraní-katalogů-otevřených-dat/2021-01-11/#položky-distribuce-balíčkovací-formát>



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost



MINISTERSTVO VNITRA
ČESKÉ REPUBLIKY

Téma

Hospodářství a finance [↗](#)

Populace a společnost [↗](#)

Vláda a veřejný sektor [↗](#)

Související geografické území

Česká republika [↗](#)

Časové pokrytí

2004-01-01 - 2004-12-31

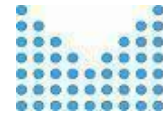
Prostorové rozlišení v metrech

20 metrů

Časové rozlišení

1 rok

Zobrazení v NKOD na úrovni distribuce datové sady bude rozšířeno následujícím způsobem:



GML

Podmínky užití distribuce	Soubor ke stažení
CC BY 4.0 Český úřad zeměměřický a katastrální	Stáhnout
CC BY 4.0 Český úřad zeměměřický a katastrální	application/gml+xml Typ média
Vlastní licence Zvláštní právo pořizovatele databáze	application/zip Formát balíčku
Neobsahuje Osobní údaje	application/zip Formát komprese

Hierarchie datových sad a datové sady bez distribucí

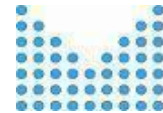
Analýza

Nové verze standardů umožňují reprezentovat vztah mezi datovými sadami, jako například že datová sada je součástí série datových sad. Tuto funkcionalitu požadují na základě zpětné vazby i poskytovatelé otevřených dat v ČR. Konkrétně se jedná o implementaci položky datové sady "je součástí"¹². S tímto souvisí i přidání možnosti registrace datové sady bez distribucí, jelikož zastřešující sada datové série žádné distribuce nemá.

Návrh

O příslušné položky bude rozšířena pipeline v LinkedPipes ETL zajišťující harverstaci lokálních katalogů a přípravu dat NKOD.

¹² <https://ofn.gov.cz/rozhraní-katalogů-otevřených-dat/2021-01-11/#položky-datová-sada-je-součástí>



Formulář pro zadávání metadat datových sad nebude rozšířen, protože tuto pokročilou metadatovou konstrukci lze rozumně podporovat pouze tam, kde je implementován Lokální katalog otevřených dat.

Zobrazení v NKOD na úrovni konkrétní datové sady v hierarchii bude rozšířeno následujícím způsobem:

Tato datová sada je součástí datové série **INSPIRE téma Územní správní jednotky (AU)**

Téma

Související geografické

Kontaktní bod

Pro zastřešující datovou sadu datové série pak na místě distribucí datové sady bude seznam datových sad v sérii takto:

Datové sady této datové série

INSPIRE - administrativní jednotky

Datová sada pro téma Územní správní jednotky (AU) harmonizovaná podle směrnice INSPIRE a datové specifikace pro ELF ve verzi 1.0. Harmonizovaná datová sada INSPIRE obsahuje pouze územní správní jednotky a jejich hranice. Územní správní jednotky jsou státy, kraje, okresy a obce. Datová sada pokrývá v jednom souboru celé území České republiky. Hranice územních samosprávných...

GML

[Zobrazit všechny dílčí datové sady v seznamu datových sad](#)

Podpora více témat datových sad

Analýza

Aktuálně umí NKOD klasifikovat datovou sadu právě jedním tématem z evropského číselníku datových témat¹³. Na základě zpětné vazby poskytovatelů je potřeba začít podporovat klasifikaci více tématy z tohoto číselníku, stejně tak tématy z jiných, uživatelských číselníků. Toto je v souladu s DCAT 2 a DCAT-AP 2.0.1.

Návrh

O vícenásobnost položky bude rozšířena pipeline v LinkedPipes ETL zajišťující harverstaci lokálních katalogů a přípravu dat NKOD.

Bude rozšířen formulář pro zadávání metadat datových sad následujícím způsobem:

Zobrazení v NKOD na úrovni datové sady bude rozšířeno následujícím způsobem:

¹³ <https://op.europa.eu/cs/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/data-theme>



Téma

Hospodářství a finance

Populace a společnost

Vláda a veřejný sektor

Podpora více položek územního pokrytí

Analýza

Nyní lze datové sadě přidat právě jeden územní prvek RÚIAN jakožto územní pokrytí. Na základě zpětné vazby od poskytovatelů a kvůli kompatibilitě s DCAT-AP je třeba toto rozšířit o další položky územního pokrytí, například pro datové sady s pokrytím mimo ČR. Toto zahrnuje práci s 3 dalšími EU číselníky pro území (stát, kontinent, místo), které jsou povinné v DCAT-AP 2.0.1.

Návrh

O příslušné položky bude rozšířena pipeline v LinkedPipes ETL zajišťující harverstaci lokálních katalogů a přípravu dat NKOD.

Bude rozšířen formulář pro zadávání metadat datových sad následujícím způsobem:

Související území

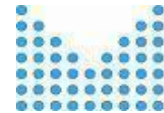
Česká republika (1) × Evropa × ?

A bude přidán dialog pro zadávání položek z jednotlivých číselníků:

Label	Value
RÚIAN	Typ územního prvku RÚIAN Území státu
KONTINENT	Název územního prvku RÚIAN Česká republika (1)
ZEMĚ	Česká republika (1)
MÍSTO	

URL JINÉHO ÚZEMÍ

PŘIDAT



Zobrazení v NKOD na úrovni datové sady bude rozšířeno následujícím způsobem:

Téma

Spravedlnost, právní systém a
veřejná bezpečnost

Doprava

Související geografické území

Brno

Moravskoslezský kraj

Podpora vícejazyčných metadat

Analýza

Na základě požadavků z řad poskytovatelů dat je třeba zajistit, aby NKOD uměl pracovat s vícejazyčnými metadaty. Tedy všechny položky, které mají povahu volného textu, bude třeba umět ve formulářích zadávat česky i anglicky, tato data umět zpracovávat a zobrazovat v uživatelském rozhraní. To zahrnuje název datové sady¹⁴, popis datové sady¹⁵, klíčová slova¹⁶, název distribuce¹⁷ a název datové služby¹⁸.

Návrh

O vícejazyčnost textových položek bude rozšířena pipeline v LinkedPipes ETL zajišťující harvestaci lokálních katalogů a přípravu dat NKOD.

Bude rozšířen formulář pro zadávání metadat datových sad následujícím způsobem pro všechny textové položky:

Název datové sady česky		Název datové sady anglicky (volitelné)	
Popis datové sady česky		Popis datové sady anglicky (volitelné)	

Zobrazení v NKOD na úrovni datové sady bude řešeno společně s přepínáním jazyka celého uživatelského rozhraní. Tedy v anglickém uživatelském rozhraní budou primárně zobrazována anglická metadata, v českém zase česká. V případě chybějící jazykové verze bude zobrazena ta, která je v metadatech přítomna.

¹⁴ <https://ofn.gov.cz/rozhraní-katalogů-otevřených-dat/2021-01-11/#položky-datová-sada-název>

¹⁵ <https://ofn.gov.cz/rozhraní-katalogů-otevřených-dat/2021-01-11/#položky-datová-sada-popis>

¹⁶ <https://ofn.gov.cz/rozhraní-katalogů-otevřených-dat/2021-01-11/#položky-datová-sada-klíčová-slova>

¹⁷ <https://ofn.gov.cz/rozhraní-katalogů-otevřených-dat/2021-01-11/#položky-distribuce-název>

¹⁸ <https://ofn.gov.cz/rozhraní-katalogů-otevřených-dat/2021-01-11/#položky-datová-slужba-název>



Podpora odkazování datové sady na Otevřené formální normy, podle kterých je publikována

Analýza

Aktuálně v NKOD není možnost u datové sady říci, dle kterých OFN je publikována, což znemožňuje jejich vyhledání právě dle implementované OFN. Je třeba tento vztah umožnit do NKOD zavést, a v NKOD ho umět zobrazit. Ve standardu DCAT-AP pro to již existuje použitelná položka, "odkaz na specifikaci"¹⁹.

Návrh

O příslušné položky bude rozšířena pipeline v LinkedPipes ETL zajišťující harverstaci lokálních katalogů a přípravu dat NKOD.

Bude rozšířen formulář pro zadávání metadat datových sad následujícím způsobem:

Specifikace (volitelné) ?

Uved'te URL všech dodržovaných otevřených formálních norem a sdílených specifikací

Zobrazení v NKOD na úrovni datové sady bude rozšířeno následujícím způsobem:

Dokumentace
Zobrazit dokumentaci

Kontaktní bod
Kontakt pro otevřená data

Specifikace

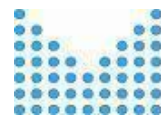
OFN: Propojená data

OFN: Číselníky

Přehledová tabulka úprav NKOD dle změn v DCAT-AP

V této tabulce jsou změny požadované na základě DCAT-AP 2.0.1 a jejich promítnutí do implementace NKOD.

¹⁹ <https://ofn.gov.cz/rozhraní-katalogů-otevřených-dat/2021-01-11/#položky-datová-sada-specifikace>



Změna	Její implementace
Změna verze z 1.2.1 na 2.0.1	Data indikující použitou verzi DCAT-AP nyní obsahují novou verzi.
Podpora pro datové služby	Zpracováno do transformačních pipeline, formuláře i zobrazení v NKOD.
Nové položky metadatového záznamu	Zpracováno do transformačních pipeline, formuláře i zobrazení v NKOD.
Hierarchie datových sad	Zpracována do transformačních pipeline a zobrazení v NKOD.
Více témat datových sad	Zpracována do transformačních pipeline, formuláře i zobrazení v NKOD.
Více položek územního pokrytí, včetně povinných evropských číselníků	Zpracována do transformačních pipeline, formuláře i zobrazení v NKOD.

Navrhované navazující aktivity MV ČR

Vzhledem k tomu, že zmiňované standardy, na kterých je NKOD postaven, se neustále vyvíjí, doporučujeme Ministerstvu vnitra nadále sledovat jejich vývoj. Aktuálně se pracuje na standardu DCAT - Version 3²⁰ a současně se pracuje na evropském doporučení DCAT-AP (2.1.0) v komunitě SEMIC²¹, na které budou navazovat práce s jeho adaptací právě na DCAT - Version 3. Tyto změny si pravděpodobně vyžádají adaptaci OFN Rozhraní katalogů otevřených dat: DCAT-AP-CZ a následné úpravy implementace v NKOD.

Přílohy

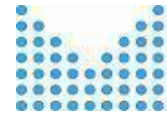
V přílohách jsou zdrojové kódy použitého open-source software a prototyp upraveného NKOD formou dokumentace a datových transformačních procesů pro nástroj LinkedPipes ETL.

Použitý open-source software

1. LinkedPipes ETL (fork v <https://github.com/opendata-mvcr/etl>)
 - o soubor `etl-develop.zip`
2. LinkedPipes DCAT-AP Viewer (fork v <https://github.com/opendata-mvcr/dcat-ap-viewer>)
 - o soubor `dcat-ap-viewer-develop.zip`

²⁰ <https://www.w3.org/TR/vocab-dcat-3/>

²¹ <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic>



3. LinkedPipes DCAT-AP Forms (fork v <https://github.com/opendata-mvcr/dcat-ap-forms>)
 - soubor `dcat-ap-forms-develop.zip`

Prototyp upraveného národního katalogu otevřených dat

1. <https://github.com/opendata-mvcr/nkod>
 - soubor `nkod-master.zip`