

Materiál Ministerstva vnitra



Export z Národní architektury eGovernmentu ČR

Obsah

Pravidla pro Veřejný datový fond 3

Pravidla pro údaje zpřístupněné veřejným datovým fondem 3

Pravidla sdílení veřejných údajů prostřednictvím VDF 4

Pravidla sdílení veřejných číselníků prostřednictvím VDF 10

Technická pravidla pro aplikační komponenty veřejného datového fondu 12

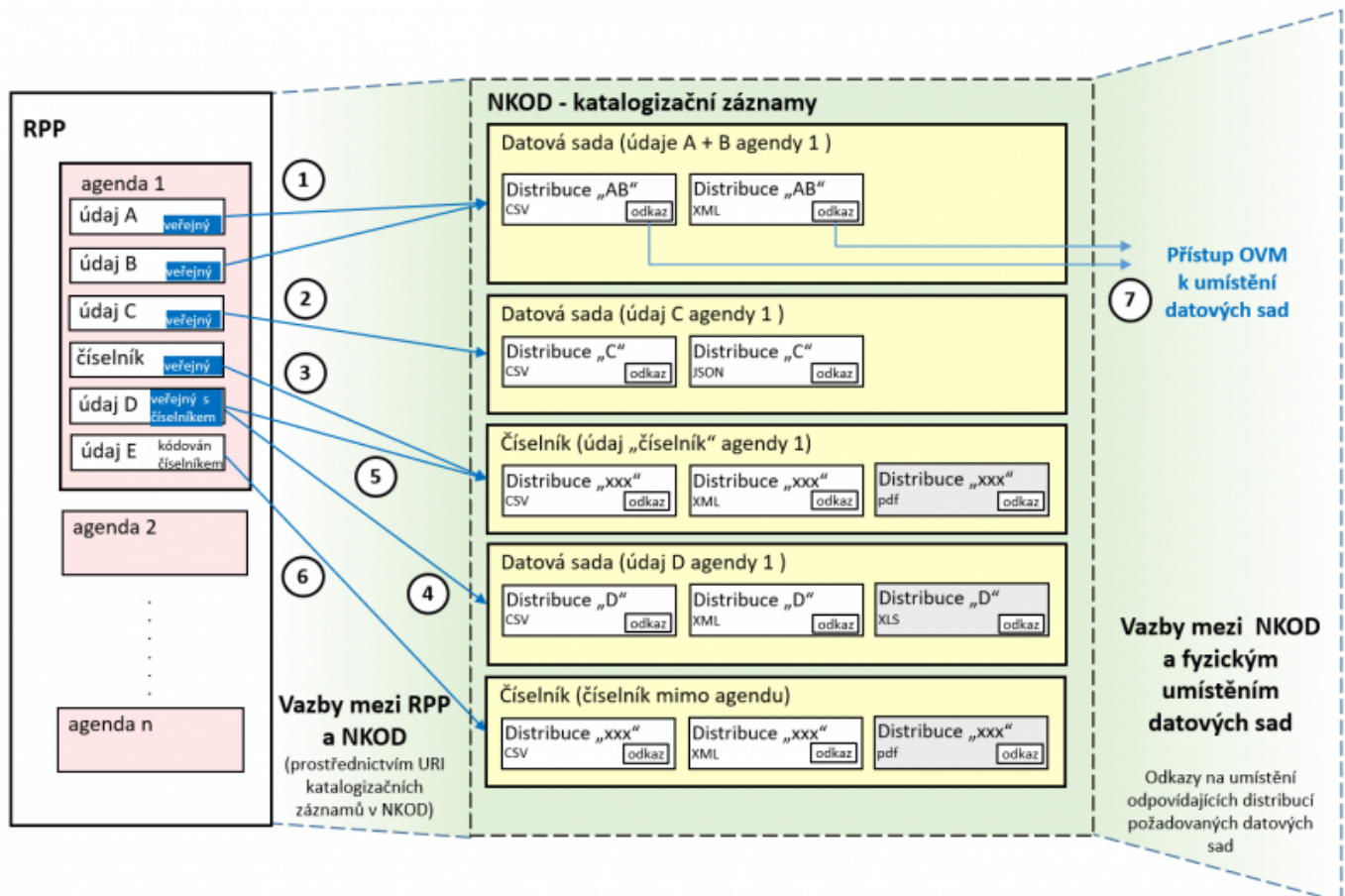
Pravidla pro Veřejný datový fond

Pravidla pro údaje zpřístupněné veřejným datovým fondem

Ve VDF jsou zpřístupňovány veřejné registrované údaje spravované jednotlivými OVM. Pro údaje zpřístupněné prostřednictvím VDF platí:

- Údaje jsou zpřístupněny v datových sadách prostřednictvím [referenčního rozhraní](#) pro potřeby čtenářů údajů - OVM a SPUÚ.
- Datové sady jsou navíc publikovány prostřednictvím otevřeného přístupu (tj. jako otevřená data dle § 3 odst. 11 [InfoZ](#)) v totožné podobě (tj. s totožnou strukturou a sémantikou).
- Otevřený přístup i přístup prostřednictvím VDF jsou tedy dva přístupy ke stejnému obsahu v podobě otevřených dat.
 - První je určen pro veřejnost, druhý je určen pro OVM a SPUÚ a je realizován prostřednictvím [referenčního rozhraní](#).
- Datové sady jsou popsány v podobě katalogizačních záznamů (metadat) v NKOD.
- Datové sady jsou fyzicky dostupné v podobě distribucí. Různé distribuce stejné datové sady zpřístupňují její obsah v různých formátech a prostřednictvím různých přístupových mechanismů. Proto je každá distribuce zaznamenána v katalogizačním záznamu datové sady v NKOD. VDF předpokládá tři následující způsoby zpřístupnění obsahu datové sady, z nichž první je povinný a zbylé dva jsou volitelné:
 - v podobě datového souboru s kompletním obsahem datové sady ke stažení,
 - v podobě API, které umožňuje přistupovat ke kompletním údajům o každé jednotlivé entitě či konceptu, o němž jsou v datové sadě reprezentovány údaje, prostřednictvím dereference identifikátoru entity či konceptu, který je stanoven poskytovatelem údajů v podobě IRI (Internationalized Resource Identifier, více viz [Otevřená formální norma pro propojená data](#)) a
 - v podobě API, které umožňuje dotazování nad obsahem datové sady s pomocí dotazovacího jazyka SPARQL
- Informace o veřejnosti registrovaného údaje je zachycena v jeho evidenci v [RPP](#) označením údaje jako veřejného údaje.
 - Pro veřejný údaj obsahuje [RPP](#) v evidenci údaje IRI datové sady (nebo datových sad) v NKOD, v níž je obsah odpovídající údají zpřístupněn prostřednictvím VDF a publikován jako otevřená data.
 - Pro údaj kódovaný číselníkem obsahuje [RPP](#) v evidenci údaje IRI datové sady v NKOD, v níž je číselník zpřístupněn prostřednictvím VDF a publikován jako otevřená data.

Vlastní mechanismus zpřístupnění údajů do VDF přibližuje dále uvedený obrázek na několika příkladech údajů „agendy 1“. Čísla v kroužcích na obrázku označují jednotlivé příklady.

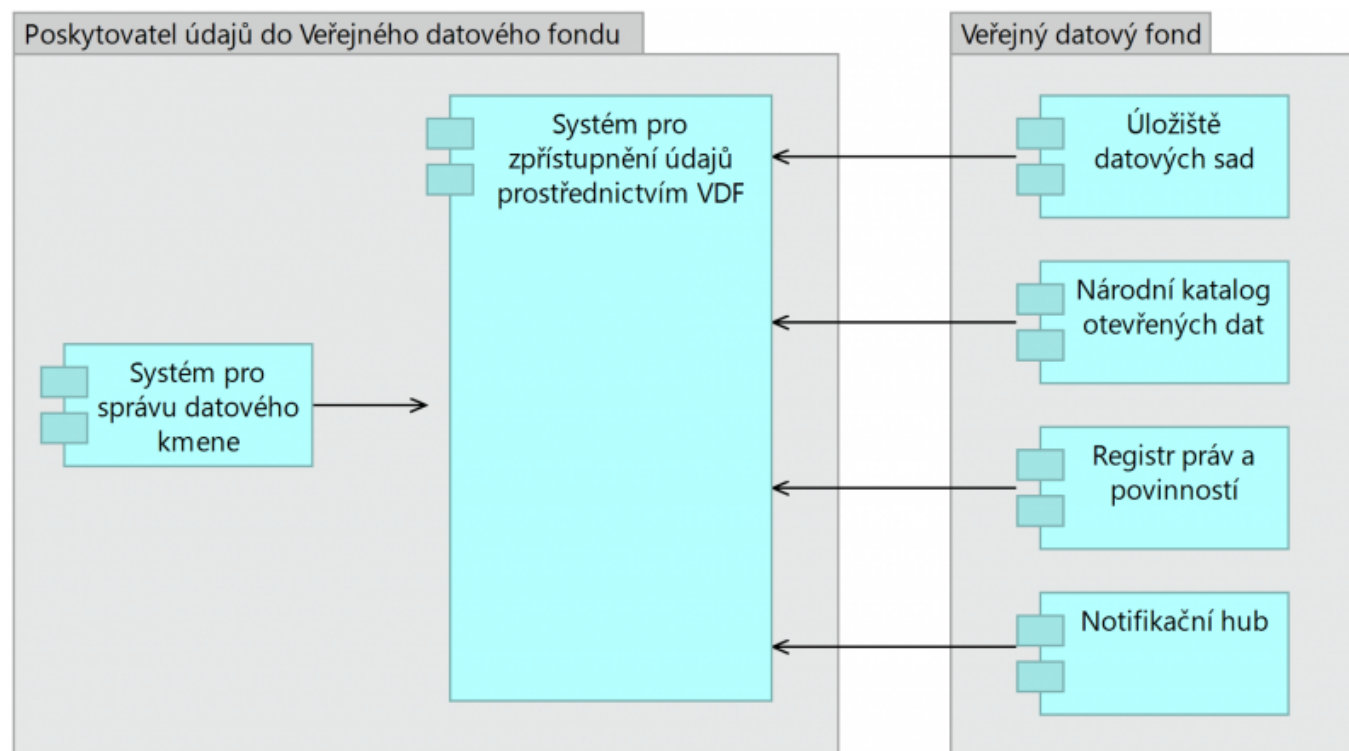


- V RPP je u agency 1 evidováno, že údaje A i B jsou veřejné. To znamená, že jsou dostupné jako otevřená data prostřednictvím VDF a otevřeného přístupu.
 - Oba údaje jsou dostupné prostřednictvím stejné datové sady.
 - V RPP je v evidenci těchto údajů uvedeno IRI datové sady v NKOD (datová sada se jmenuje „Údaje A + B agency 1“).
 - Datová sada je publikovaná a dostupná v několika distribucích, katalogizační záznam obsahuje pro každou distribuci odkaz na její fyzické umístění (tj. její URL).
- V RPP je u agency 1 evidováno, že údaj C je veřejný. Jedná se o stejnou situaci jako v příkladu 1, pouze s tím rozdílem, že údaj C je publikován v jiné samostatné datové sadě.
- Agenda 1 vytváří a udržuje číselník, který je dostupný ve VDF a je publikován jako otevřená data.
- V RPP je u agency 1 evidováno, že údaj D je veřejný a je publikován v samostatné datové sadě. Jedná se o stejnou situaci jako v příkladu 1.
- U údaje D je v RPP evidováno, že je kódován číselníkem (v příkladu je uvedena situace s číselníkem, který agenda přímo vytváří, ale vše uvedené platí, pro jakýkoliv využívaný číselník). Evidence údaje v RPP proto také obsahuje IRI datové sady v NKOD, který obsahuje publikovaný číselník.
- V RPP je u agency 1 evidováno, že údaj E je neveřejný a tudíž není možné jej zpřístupnit ve VDF. U údaje E je ale evidováno, že je kódován číselníkem (v tomto případě se pro demonstraci jedná o číselník spravovaný mimo agendu 1). Evidence údaje v RPP obsahuje IRI datové sady v NKOD, který obsahuje publikovaný číselník (stejně jako v příkladu 5).
- Reprezentuje přístup k distribucím datových sad prostřednictvím VDF.

Pravidla sdílení veřejných údajů prostřednictvím VDF

Pravidla publikace veřejných údajů do VDF

Základní prvky architektury VDF z pohledu poskytovatele údajů zobrazuje následující obrázek.



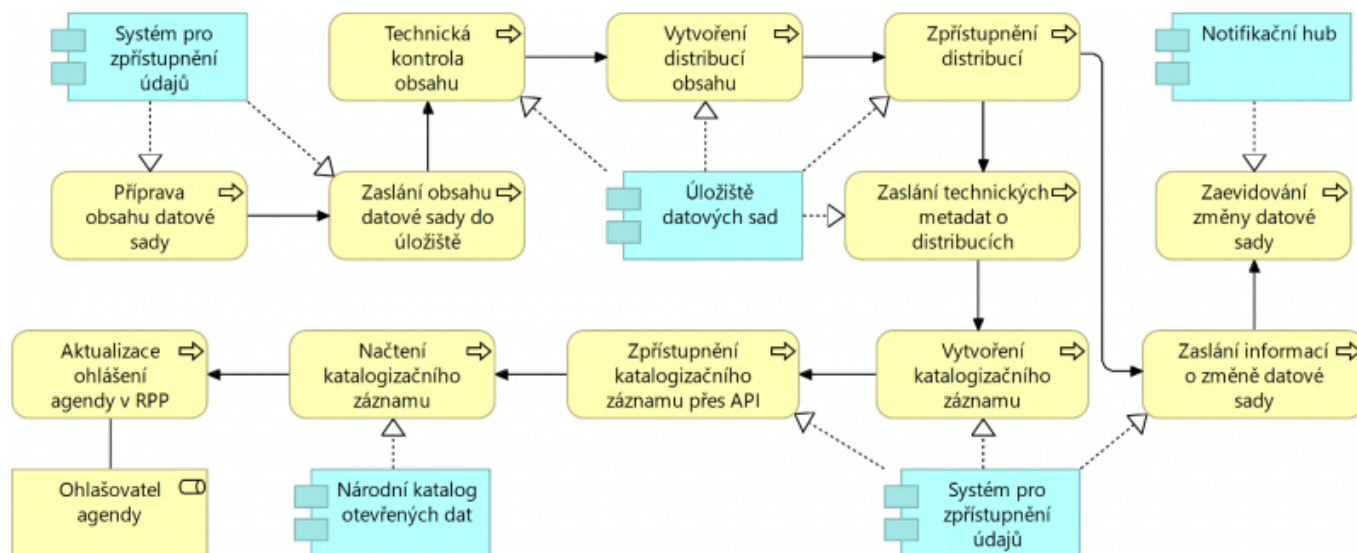
Poskytovatelem údajů do VDF je správce ISVS, ve kterém jsou vedeny registrované veřejné údaje. Tento ISVS je vyznačen na levé straně obrázku jako *systém pro správu datového kmene*, kterým OVM spravuje svůj datový kmen. V praxi se samozřejmě může jednat o více ISVS, zde si pro jednoduchost zobrazujeme jen jeden systém.

Pro potřeby sdílení údajů ve VDF poskytovatel údajů vytvoří *systém pro zpřístupnění údajů prostřednictvím VDF*. Může se jednat o samostatný systém nebo to může být modul v rámci existujícího systému. Zajišťuje získávání obsahu veřejných údajů z datového kmene poskytovatele, rozdělení do vhodných datových sad a převod do podoby definované otevřenými formálními normami a jeho dávkové předání do úložiště datových sad.

Úložiště datových sad zajišťuje kontrolu technické správnosti zaslaných dávek vůči otevřeným formálním normám a zpřístupnění distribucí obsahu ve všech formátech definovaných otevřenými formálními normami. Úložiště dále zajišťuje dostupnost distribucí čtenářům prostřednictvím ISGOD i prostřednictvím veřejného internetu. Dostupnost distribucí prostřednictvím ISGOD navíc garantuje. Úložiště datových sad pro ukládání obsahu datových sad z daného ISVS je vytvořeno pro daný ISVS právě jedno a spravuje jej správce ISVS. V případě, že se jedná o aktualizaci obsahu datové sady, oznamuje úložiště datových sad notifikačnímu hubu, že obsah datové sady byl změněn.

Poté, co jsou distribuce obsahu datových sad uloženy v úložišti a zpřístupněny, jsou datové sady katalogizovány v NKOD prostřednictvím systému pro zpřístupnění údajů. K tomu poskytuje systém pro zpřístupnění údajů API, které splňuje [otevřenou formální normu rozhraní katalogů otevřených dat](#). Katalogizace datových sad v NKOD je tak automatizovaná.

Celý proces publikace údajů sdružených v jedné datové sadě do VDF je znázorněn na následujícím diagramu. Proces předpokládá, že příslušná agenda již byla ohlášena v [RPP](#) včetně všech jejích údajů v potřebné úrovni granularity.



V rámci procesu:

- Systém pro zpřístupnění údajů
 - Připraví obsah datové sady v podobě datového souboru v jednom z formátů definovaných otevřenými formálními normami.
 - Specifikaci otevřených formálních norem lze získat z repozitáře otevřených formálních norem.
 - Pokud pro údaje neexistuje otevřená formální norma, musí ji správce systému pro zpřístupnění údajů s podporou MV ČR nejprve vytvořit.
 - Záže připravený obsah datové sady do úložiště datových sad.
- Úložiště datových sad
 - Provede technickou kontrolu zaslání obsahu
 - Kontrola správného formátování (např. JSON nebo XML formátování)
 - Kontrola validity datové struktury vůči datovým schémátům definovaných otevřenými formálními normami (např. vůči JSON nebo XML schémátům)
 - V případě špatné syntaxe záže zpět systému pro zpřístupnění údajů chybové hlášení a skončí.
 - Vytvoří distribuce obsahu jeho transformací do všech podob definovaných otevřenými formálními normami s využitím transformačních skriptů/procedur/mapování, které jsou součástí otevřených formálních norem.
 - Zpřístupní vytvořené distribuce
 - Zpřístupní je jako datové soubory dostupné ke stažení prostřednictvím ISGOD a z veřejného internetu.
 - URL pro stažení datového souboru je stejné pro přístup prostřednictvím ISGOD a veřejného internetu, k čemuž je nutné správně nastavit DNS v prostředí KIVS/CMS a DNS v prostředí veřejného internetu.
 - Volitelně zpřístupní jednotlivé položky obsahu dle [otevřené formální normy pro propojená data](#) tak, že má každá položka své referenční a lokální IRI dereferencovatelné prostřednictvím ISGOD a z veřejného internetu.
 - Referenční IRI položky je stejné pro přístup prostřednictvím ISGOD a veřejného internetu, k čemuž je nutné správně nastavit DNS v prostředí KIVS/CMS a DNS v prostředí veřejného internetu.
 - Lokální IRI položky je stejné pro přístup prostřednictvím ISGOD a veřejného internetu, k čemuž je nutné správně nastavit DNS v prostředí KIVS/CMS a DNS v prostředí veřejného internetu.
 - Volitelně zpřístupní jejich obsah v podobě SPARQL endpointu prostřednictvím ISGOD a veřejného internetu.
 - URL SPARQL endpointu je stejné pro přístup prostřednictvím ISGOD a veřejného internetu, k čemuž je nutné správně nastavit DNS v prostředí KIVS/CMS a DNS v prostředí veřejného internetu.

- Zašle zpět systému pro zpřístupnění údajů potvrzení o úspěšném uložení.
 - Jako součást potvrzení zasílá metadata o vytvořených distribucích v [podobě definované otevřenou formální normou pro rozhraní katalogů otevřených dat](#).
- Systém pro zpřístupnění údajů
 - Vytvoří kompletní katalogizační záznam o datové sadě včetně metadat o distribucích vytvořených úložištěm datových sad a zpřístupní jej prostřednictvím API dle [otevřené formální normy pro rozhraní katalogů otevřených dat](#).
 - Zašle notifikačnímu hubu informaci o změně obsahu datové sady.
 - Úroveň detailu informace není v tomto místě řešena.
- Národní katalog otevřených dat
 - Získá katalogizační záznam z API poskytnutého systémem pro zpřístupnění údajů a zaeviduje jej.
- Ohlašovatel agendy
 - Ohlásí do [RPP](#) jako součást ohlášení agendy referenční IRI datové sady (datových sad) v NKOD, ve které (kterých) je veřejný údaj zpřístupněn. Ohlášení provede poté, co NKOD datovou sadu na základě zaslání katalogizačního záznamu zaeviduje (zpravidla do 1 dne).
- Notifikační hub
 - Zaeviduje informaci o změně datové sady zaslanoú úložištěm datových sad.

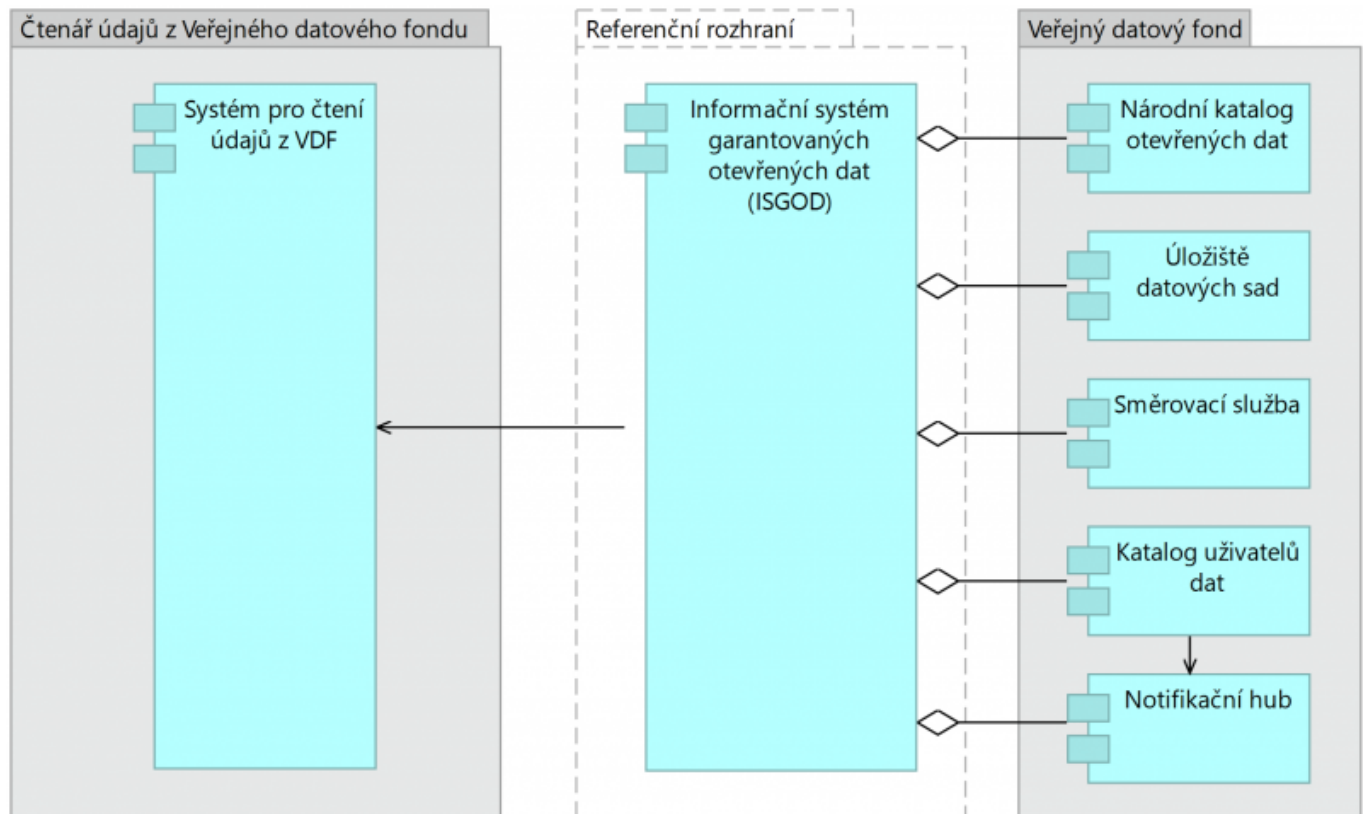
Systém pro zpřístupnění údajů prostřednictvím VDF souvisí s existujícím systémem pro zpřístupnění údajů prostřednictvím [PPDF](#), který zajišťuje poskytování údajů o konkrétním subjektu práva, na který přistupuje čtenářský AIS prostřednictvím [PPDF](#). Systém pro zpřístupnění údajů prostřednictvím VDF (dále jen *systém pro zpřístupnění údajů*) oproti tomu aktivně v pravidelných správcem ISVS definovaných intervalech exportuje obsah veřejných údajů do podoby datových sad a dávkově je předává do *úložiště datových sad*, ze kterého jsou jako otevřená data dostupná prostřednictvím VDF a otevřeného přístupu.

[PPDF](#) a VDF jsou tedy dva různé způsoby sdílení datového kmene agendy. Technická podoba dat určená pro sdílení prostřednictvím [PPDF](#) je definována v kontextech, tj. XSD schématech popisujících XML struktury, ve kterých je obsah datového kmene agendy sdílen prostřednictvím [PPDF](#). Technická podoba dat určená pro sdílení prostřednictvím VDF je definována otevřenými formálními normami. Otevřené formální normy definují datová schémata. Nejedná se ale nutně jen o XSD schémata, ale také o JSON schémata, CSV schémata nebo ontologie pro popis RDF reprezentace. To z toho důvodu, že totožný obsah, který je dostupný prostřednictvím VDF je dostupný jako otevřená data, kde je nutno z důvodů interoperability a dodržení dobré praxe nabídnout obsah v různých standardních formátech.

Protože ale kontexty pro [PPDF](#) a datové struktury v OFN pro VDF jsou dvěma syntaktickými stranami téže sémantické mince (tj. jsou různými syntaktickými reprezentacemi stejné sémantiky), je nutno tuto sémantiku strukturovaně a explicitně vyjádřit. K tomu jsou využívány techniky ontologického konceptuálního modelování, kdy je sémantika všech údajů v dané agendě popsána na konceptuální úrovni v podobě ontologie podle vyhlášky, která nahradí současnou vyhlášku č. 529. Pro tvorbu konceptuálních modelů v podobě ontologie MV ČR spravuje a provozuje sadu volně dostupných modelovacích nástrojů. Ty umožňují také z konceptuálních modelů definice kontextů pro [PPDF](#) a datových struktur v OFN pro VDF automatizovaně generovat a zajišťovat tak jejich vzájemnou sémantickou interoperabilitu. Konceptuální model agendy by navíc měl být tvořen konzistentně s modely ostatních agend, a modely agend by měly vycházet ze společné ontologie veřejné správy a ze slovníků definovaných EU (tzv. ISA Core Vocabularies), což podporuje sémantickou interoperabilitu vyměňovaných údajů napříč agendami i v rámci EU.

Čtení veřejných údajů z VDF

Základní stavební kameny architektury VDF z pohledu čtenáře údajů zobrazuje následující obrázek.



Čtenářem údajů z VDF je správce ISVS, který čte veřejné údaje. Tento ISVS je v obecné úrovni vyznačen na levé straně obrázku jako *systém pro čtení údajů z VDF* (dále jen *systém pro čtení údajů*).

Systém pro čtení údajů čte veřejné údaje z VDF jako otevřená data prostřednictvím ISGOD v podobě distribucí datových sad v různých formátech definovaných otevřenými formálními normami. Jsou umožněny 3 základní druhy přístupu prostřednictvím ISGOD:

1. Přístup ke kompletnímu obsahu datové sady v podobě datových souborů voláním
 1. služeb ISGOD umožňujících přistoupit k metadatům o datové sadě a jejich distribucích na základě jejich referenčních IRI a k URL daného souboru a stáhnout jej. (povinné)
2. Přístup k jednotlivým položkám datových sad voláním služeb ISGOD umožňujících přistoupit k datům o dané položce na základě jejího referenčního IRI. (volitelné)
3. Dotazování nad položkami datových sad voláním dotazovacích služeb ISGOD. (volitelné)

Služby ISGOD jsou realizovány jako webové služby postavené na principech REST, které jsou poskytovány jednotlivými komponentami VDF znázorněnými v pravé části obrázku:

- REST služby NKOD umožňují číst metadata o datových sadách a jejich distribucích.
- REST služby úložiště datových sad umožňují číst obsah v nich uložených datových sad v podobě
 - stahování datových souborů s obsahem uložených datových sad (povinné)
 - přístupu k IRI jednotlivých položek obsahu uložených datových sad (volitelné)
 - SPARQL dotazů nad obsahem uložených datových sad (volitelné)

ISGOD je pouhým logickým zastřešením výše uvedených služeb.

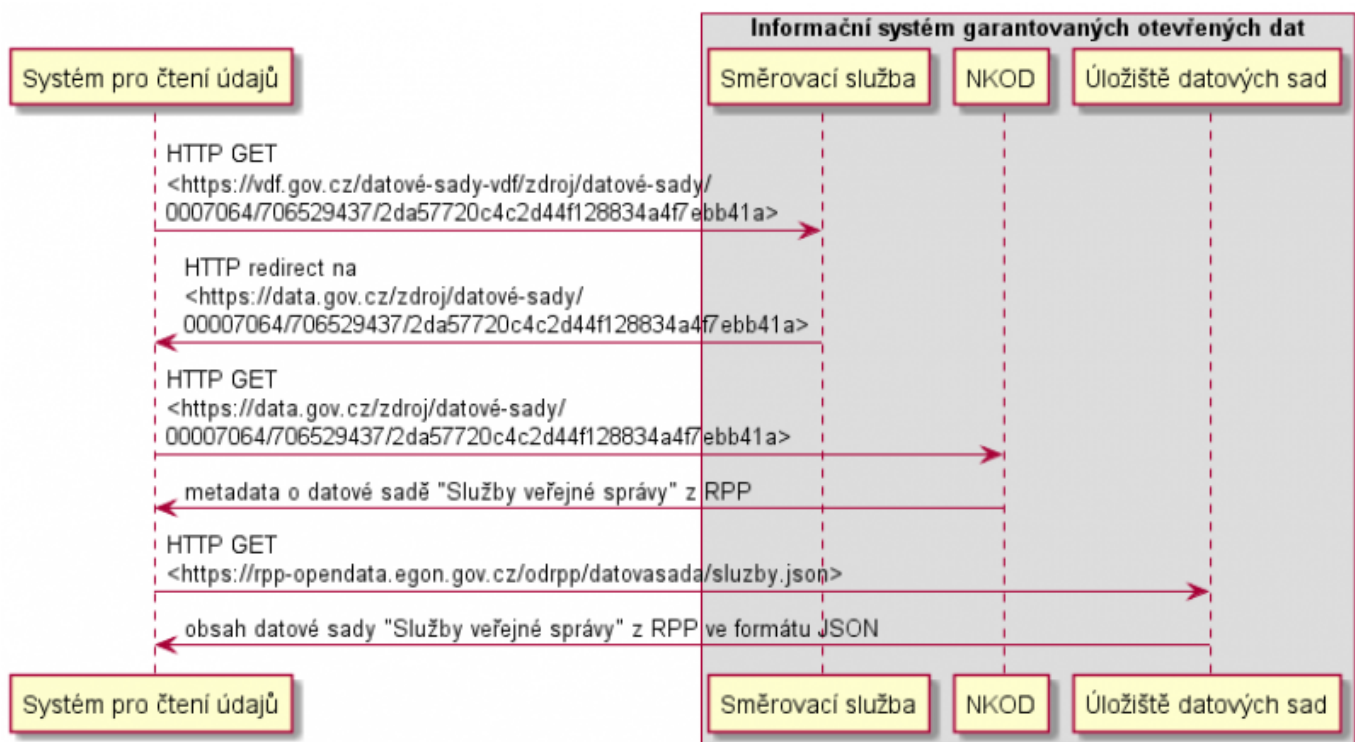
Čtení obsahu datové sady v podobě datového souboru

Čtení obsahu datové sady v podobě datového souboru typicky systém pro čtení údajů provádí za účelem aktualizace vlastní kopie údajů přebíraných z VDF. Aktualizaci typicky provádí v pravidelných intervalech nebo na základě notifikací o změnách zasílaných notifikačním hubem na základě registrace v katalogu uživatelů, ale mimo svůj run-time. Dávkový přístup ke kompletnímu obsahu datové sady v podobě datového souboru

předpokládá, že systém pro čtení údajů zná referenční IRI datové sady. Referenční IRI datové sady je možné zjistit z evidence agendových údajů v [RPP](#) nebo vyhledáváním v NKOD. Přístup je pak realizován následujícím postupem:

- Systém pro čtení údajů přistupuje k referenčnímu IRI datové sady.
- Směrovací služba přeměňuje referenční IRI datové sady na lokální IRI datové sady v NKOD.
- Systém pro čtení údajů přistupuje k lokálnímu IRI datové sady v NKOD.
- NKOD vrací metadata o datové sadě.
- Systém pro čtení údajů vybírá distribuci datové sady dle potřebného formátu a přistupuje k URL ke stažení obsahu distribuce.
- Úložiště datových sad zasílá systému pro čtení údajů obsah datového souboru na daném URL.

Následující obrázek postup znázorňuje v podobě sekvenčního UML diagramu na konkrétním příkladu přístupu k datové sadě "Služby veřejné správy", která je publikována z [RPP](#).



Čtení položky datové sady

Čtení položky datové sady typicky systém pro čtení údajů provádí za účelem zobrazení veřejných údajů o položce v uživatelském rozhraní nebo jiné práce s konkrétní položkou v okamžiku potřeby práce s údají o položce, tj. v rámci svého run-time. Přístup k položce předpokládá, že systém pro čtení údajů zná referenční IRI položky. Referenční IRI položky je možné získat následujícími způsoby:

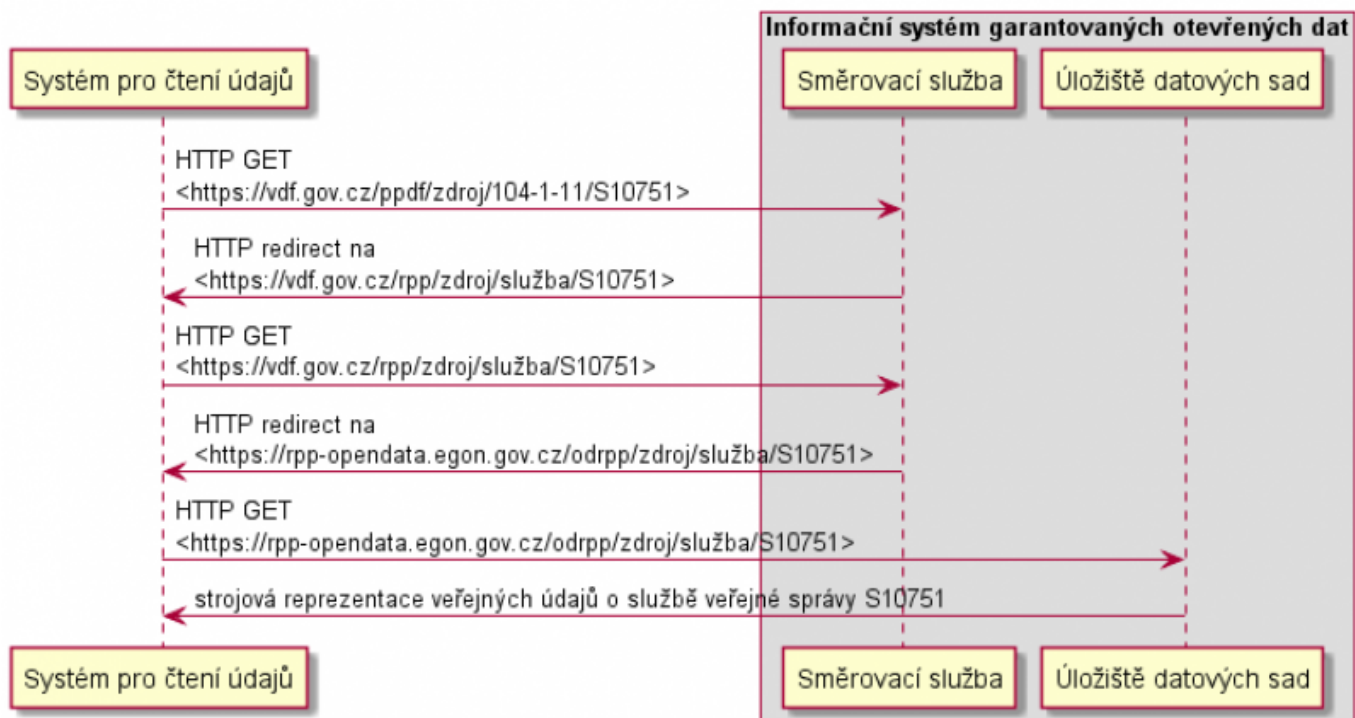
- V předchozích krocích byl přečten z VDF údaj s referenčním IRI jako hodnotou.
- V předchozích krocích byl přečten z [PPDF](#) údaj s proprietárním identifikátorem entity jako hodnotou (tj. identifikátor v podobě řetězce, který identifikuje entitu lokálně v rámci ISVS). Referenční IRI položky s veřejnými údaji o entitě získá systém pro čtení údajů voláním směrovací služby s kódem údaje (evidovaným v [RPP](#)) a proprietárním identifikátorem.

Přístup je pak realizován následujícím postupem:

- Systém pro čtení údajů přistupuje k referenčnímu IRI položky.
- Směrovací služba přeměňuje referenční IRI položky na lokální IRI položky v konkrétním úložišti datových sad, kde jsou údaje o položce uloženy.

- Systém pro čtení údajů přistupuje k lokálnímu IRI položky na daném úložišti datových sad.
- Úložiště datových sad vrací veřejné údaje o položce.
- Systém pro čtení údajů zobrazuje nebo jinak zpracovává získané údaje.

Následující obrázek postup znázorňuje v podobě sekvenčního UML diagramu na konkrétním příkladu přístupu k veřejným údajům o službě evidované v RPP s proprietárním identifikátorem S10751 a s názvem "Pěstitelské pálení". Jedná se o položku datové sady "Služby veřejné správy", která je publikována z RPP. Postup zahrnuje i získání referenčního identifikátoru položky z proprietárního identifikátoru na začátku procesu. Systému pro čtení údajů je známo pouze id "S10751" pro údaj agendy 104 s kódem "104-1-11". Zkonstruuje IRI identifikující položku a přistoupí na něj. Toto IRI vede na směrovací službu, která provede přesměrování na referenční IRI položky.



Pravidla sdílení veřejných číselníků prostřednictvím VDF

Speciálním případem sdílení veřejných údajů prostřednictvím VDF je pak sdílení veřejných číselníků. Vychází z architektury sdílení veřejných údajů popsané v předchozí kapitole. Má však svá specifika, která jsou popsána zde.

Pravidla publikace veřejných číselníků do VDF

Architektura pro publikaci veřejných číselníků do VDF je vystavěna na bázi architektury pro publikaci veřejných údajů do VDF. Aby byl veřejný číselník publikován do VDF, musí být podle § 51 odst. 8 ZoZr zaveden do RPP. Zavedení je provedeno prostřednictvím AIS působnostní a provede jej buď ohlašovatel agendy nebo ČSÚ (dále dohromady jen *poskytovatel číselníku*). ČSÚ zavádí do RPP veřejné číselníky nezávisle na agendách. Ohlašovatel agendy zavádí veřejný číselník jen v případě, že je agendový údaj kódován číselníkem, který ještě není v RPP nikým zaveden.

Všechny veřejné číselníky jsou tedy jako referenční údaje evidovány v RPP prostřednictvím AIS působnostní a z něj jsou také publikovány do VDF. Z AIS působnostní jsou také publikovány do VDF všechny ostatní veřejné údaje evidované v RPP v podobě datových sad. Z pohledu [architektury pro publikaci veřejných údajů do VDF](#) je tedy pro potřeby všech veřejných číselníků AIS působnostní systémem pro správu datového kmene a zároveň

má jako svoji komponentu systém pro zpřístupnění údajů prostřednictvím VDF, který zajišťuje publikaci obsahu veřejných číselníků do VDF. Pro ukládání obsahu veřejných číselníků a také obsahu veřejných údajů vedených v [RPP](#) je využito stávající úložiště, na kterém je uložen obsah [RPP](#) publikovaný jako otevřená data. AIS působnostní zajišťuje také API poskytující katalogizační záznamy o jednotlivých datových sadách s veřejnými číselníky a s obsahem údajů vedených v [RPP](#).

K realizaci výše popsané architektury publikace veřejných číselníků do VDF a jako otevřená data je nutno zajistit následující rozšíření informačního systému AIS působnostní a [RPP](#):

1. označování veřejnosti a neveřejnosti údaje
 1. včetně odkazů na legislativu v případě neveřejnosti údaje
 2. včetně IRI datových sad v NKOD, prostřednictvím kterých je veřejný údaj publikován
2. evidenci veřejných číselníků
 1. pro každý veřejný číselník existuje 1..- verzí, které chápeme jako jednotlivé datové sady
 2. všechny datové sady reprezentující jednotlivé verze číselníku jsou seskupeny do zastřešující datové sady
 3. pro zastřešující datovou sadu a jednotlivé verze jsou evidována metadata [datové sady](#) dle [otevřené formální normy pro rozhraní katalogů otevřených dat](#)
 1. mimo vlastnosti [poskytovatel](#), protože tato vlastnost reprezentuje poskytovatele datové sady do VDF
 1. kterým je u číselníků vždy MV ČR, nikoliv poskytovatel číselníku
 2. pro zastřešující datovou sadu je navíc evidováno
 1. OVM, který zavádí veřejný číselník do [RPP](#), jako poskytovatele číselníku
 1. což není poskytovatel datové sady s číselníkem do VDF, kterým je v případě veřejných číselníků vždy MV ČR, viz předchozí bod
 3. pro verzi číselníku je navíc evidováno
 1. lokální proprietární identifikátor či kód číselníku
 1. potřebné pro konstrukci lokálních IRI číselníků a jejich položek
 2. může vyplnit poskytovatel číselníku nebo je vygenerováno automaticky, pokud poskytovatel číselníků vlastní identifikátor či kód číselníku neeviduje
 4. pro datové sady reprezentující jednotlivé verze číselníku jsou navíc evidovány následující vazby, které nejsou evidovány pro zastřešující datovou sadu:
 1. Je verzí (reference na zastřešující datovou sadu)
 2. Má předchozí verzi (reference na datovou sadu s předchozí verzí číselníku, existuje-li)
 3. zavedení nového veřejného číselníku poskytovatelem číselníku
 1. poskytovatel číselníku specifikuje metadata pro zastřešující datovou sadu číselníku
 1. lze převzít nebo jinak použít [existující formulář pro registraci datové sady](#)
 2. poskytovatel číselníku specifikuje metadata pro datovou sadu s první verzí číselníku
 1. může zvolit možnost kopírovat hodnoty zadané pro zastřešující datovou sadu
 3. poskytovatel číselníku předá obsah první verze číselníku ručně v uživatelském rozhraní nahráním připraveného souboru s obsahem první verze veřejného číselníku v podobě definované otevřenou formální normou
 4. předchozí tři body lze realizovat také automatizovaně načtením seznamu veřejných číselníků poskytovatele z URL, které zadá
 1. seznam musí být zpřístupněn dle [otevřené formální normy pro rozhraní katalogů otevřených dat](#).
 2. veřejné číselníky ale nemusí být pro účely předání zpřístupněny jejich správcem jako otevřená data.
 5. předaný obsah je zvalidován vůči otevřené formální normě pro číselníky
 6. obsah je uložen v podobě zkontrolovaného předaného datového souboru
 1. obsah veřejného číselníku pouze eviduje, ale nejsou nad ním stavěny žádné aplikační funkce
 4. zavedení nové verze již zavedeného veřejného číselníku poskytovatelem číselníku
 1. stejný postup jako při zavádění nového veřejného číselníku, ale je zavedena pouze další verze číselníku zařazená pod zastřešující datovou sadu
 2. původní verze zůstává evidována včetně její publikace do VDF a jako otevřená data
 5. funkcionality systému pro zpřístupnění údajů prostřednictvím VDF

1. veřejné číselníky již jsou evidovány v podobě souborů s jejich jednotlivými verzemi v podobě definované otevřenými formálními normami, čili je nutno pouze zajistit jejich předání do úložiště veřejných číselníků a datových sad [RPP](#)
2. další veřejné údaje evidované v [RPP](#) již jsou získávány z interní databáze [RPP](#) a AIS působnostní v podobě definované otevřenými formálními normami a předávány do úložiště veřejných číselníků a datových sad [RPP](#)
6. funkcionality úložiště veřejných číselníků a datových sad [RPP](#)
 1. bude vytvořeno ze stávajícího úložiště obsahu datových sad publikovaných z [RPP](#) jako otevřená data
 2. jako doposud bude zpřístupňovat obsah [RPP](#) jako datové sady dle příslušných [otevřených formálních norem](#) ve formátech JSON, JSON-LD a prostřednictvím SPARQL endpointu
 3. zajistí také publikaci distribucí datových sad s verzemi veřejných číselníků evidovaných v [RPP](#) dle otevřené formální normy pro číselníky (formáty XML, CSV, JSON-LD a SPARQL endpoint)
 4. jelikož se jedná pouze o komponentu v rámci AIS působnostní, resp. [RPP](#), není nutné zajišťovat všechny funkcionality přesně podle obecné architektury publikace veřejných údajů do VDF
 5. je nutné zajistit dostupnost nejen z veřejného internetu jako doposud, ale také prostřednictvím ISGOD ([referenční rozhraní](#)) a garantovat dostupnost
7. funkcionality lokálního katalogu otevřených dat pro katalogizaci datových sad publikovaných v úložišti veřejných číselníků a datových sad [RPP](#)
 1. zpřístupňuje do NKOD katalogizační záznam pro každou datovou sadu:
 1. datové sady zastřešující verze číselníků a datové sady s verzemi číselníků
 1. metadata o datových sadách jsou získány od poskytovatele
 2. metadata o distribucích jsou doplněny automatizovaně na základě vytvářených distribucí v úložišti
 2. datové sady s obsahem dalších veřejných údajů evidovaných v [RPP](#) (tj. ty, které jsou již dnes publikovány jako otevřená data)
 1. metadata jsou fixně předvyplněna
 2. je registrován pod MV ČR
8. označování údaje jako údaje kódovaného verzí veřejného číselníku
 1. včetně zaznamenávání IRI datové sady s touto verzí veřejného číselníku z NKOD
 1. Aby mohlo být IRI zaznamenáno, musí být daná verze veřejného číselníku nejprve do [RPP](#) zavedena, publikována do VDF a katalogizována v NKOD.
9. evidence veřejných údajů využívaných ohlášenou agendou

Kromě nových verzí veřejného číselníku existuje možnost, že je číselník kompletně nahrazen zcela novým číselníkem. V tom případě je skutečně zaveden jako zcela nový číselník bez vazby na původní číselník. Původní číselník ale zůstává evidován.

Číselníky kódující údaje evidované v [RPP](#) již nebudou publikovány jako otevřená data stávajícím mechanismem. Stanou se veřejnými číselníky, tj. budou evidovány v AIS působnostní a budou z něj publikovány do VDF a jako otevřená data standardním výše popsaným způsobem.

Pravidla čtení veřejných číselníků z VDF

Čtení veřejných číselníků včetně jejich obsahu jako celku v podobě datových souborů ke stažení (VDF), přístupu k jednotlivým položkám datových sad s verzemi číselníků (VDF a otevřená data) a dotazování prostřednictvím SPARQL endpointu (otevřená data) probíhá v rámci [architektury pro čtení veřejných údajů popsané výše](#). Veřejné číselníky jsou dostupné dle otevřené formální normy pro číselníky.

Technická pravidla pro aplikační komponenty veřejného datového fondu

Úložiště datových sad

Úložiště datových sad je složeno ze 3 modulů:

- souborové úložiště distribucí datových sad
 - ukládá distribuce v podobě datových souborů
 - zpřístupňuje datové soubory distribucí prostřednictvím VDF a veřejného internetu
 - každý datový soubor je dostupný na jednom URL, které je stejné pro VDF i veřejný internet
 - nutno správně nastavit DNS pro KIVS/CMS a DNS pro veřejný internet
- modul pro validaci a transformaci distribucí dle příslušných otevřených formálních norem
 - kontroluje správné formátování a validitu
 - provádí transformace mezi jednotlivými formáty s využitím definic transformací v otevřených formálních normách
 - ukládá výsledky transformací do souborového úložiště a v případě RDF distribucí také do triplestore
 - bez definované otevřené formální normy pro daný typ dat není možné údaje prostřednictvím VDF zpřístupňovat
 - příslušnou otevřenou formální normu nebo normy získává modul z repozitáře otevřených formálních norem
- triplestore pro ukládání RDF distribucí
 - ukládá RDF distribuce datové sady dle otevřených formálních norem v triplestore (triplestore = databázový systém pro ukládání RDF dat v podobě trojic)
 - zpřístupňuje SPARQL endpoint pro dotazování nad RDF reprezentací a HTTP dereferenci IRI položek prostřednictvím rozhraní pro čtení distribucí jako otevřená data
 - lokální IRI položky je stejné pro VDF i veřejný internet, URL SPARQL endpointu stejné pro VDF i veřejný internet
 - nutno správně nastavit DNS pro KIVS/CMS a DNS pro veřejný internet

Směrovací služba pro veřejné číselníky a jejich položky

Jak bylo popsáno výše veřejný číselník bude podle otevřené formální normy pro číselníky zpřístupněn jako datový soubor ke stažení a prostřednictvím dereference IRI jednotlivých položek. Je tedy nutno určit tvar referenčních a lokálních IRI položek veřejných číselníků a také samotných číselníků. Ta jsou určena dle [pravidel pro tvorbu IRI](#) následovně:

- Referenční IRI číselníku:

`https://vdf.gov.cz/číselníky-vdf/zdroj/číselníky/<ID číselníku v RPP>`

- Referenční IRI verze číselníku k DDDD-MM-YY

`https://vdf.gov.cz/číselníky-vdf/zdroj/číselníky/<ID číselníku v RPP>/<DDDD-MM-YY>`

- Lokální IRI číselníku:

`https://rpp-opendata.egon.gov.cz/odrpp/zdroj/číselníky/<ID číselníku v RPP>`

- Lokální IRI verze číselníku:

`https://rpp-opendata.egon.gov.cz/odrpp/zdroj/číselníky/<ID číselníku v RPP>/<DDDD-MM-YY>`

- Referenční IRI položky číselníku:

`<referenční IRI číselníku>/položky/<lokální kód položky>`

- Lokální IRI položky číselníku:

<lokální IRI číselníku>/položky/<lokální kód položky>

Kde

- <ID číselníku v RPP> značí neměnné veřejné ID identifikující číselník v RPP
- <DDDD-MM-YY> značí datum vydání verze číselníku
- <lokální kód položky> značí kód položky číselníku v rámci daného číselníku

Pro potřeby sdílení veřejných číselníků je ve směrovací službě směrování výše uvedených referenčních IRI na lokální IRI přednastaveno a není potřeba, aby správce **RPP** nebo poskytovatelé jednotlivých číselníků směrování konfigurovali.

Dále je potřeba ve směrovací službě nastavit směrování na referenční IRI pro případy, kdy je znám pouze <lokální kód položky> a **RPP** identifikátor agendového údaje, jehož je hodnotou. Konfiguraci tohoto směrování provádí **RPP**. Kdykoliv ohlašovatel agendy uvede pro daný agendový údaj veřejný číselník, má **RPP** evidováno, kdo je poskytovatelem číselníku. Do směrovací služby tedy zaeviduje pravidlo pro směrování dvojice

(<RPP identifikátor agendového údaje>, <lokální kód položky>)

na referenční IRI

<https://vdf.gov.cz/číselníky-vdf/zdroj/číselníky/<ID číselníku v RPP>/položky/<lokální kód položky>>

From:

<https://archi.gov.cz/> - **Architektura eGovernmentu ČR**

Permanent link:

https://archi.gov.cz/nap:pravidla:pravidla_veřejny_datovy_fond

Last update: **2021/04/30 10:47**

