

Materiál Ministerstva vnitra



Export z Národní architektury eGovernmentu ČR

Obsah

Veřejný datový fond 3
 Popis veřejného datového fondu 3
 Pravidla veřejného datového fondu 13
 Příloha 1: Metodika poskytování a čerpání údajů prostřednictvím VDF 24
 Příloha 2: Srovnání vlastností jednotlivých způsobů přístupu 24

Veřejný datový fond

Popis veřejného datového fondu

Veřejný datový fond (VDF) je definován v [Informační koncepci ČR \(IKČR\)](#) jako [dílčí cíl 5.10](#) a je součástí budovaného eGovernmentu VS ČR.

“Veřejný datový fond tvořený publikovanými veřejnými údaji veřejné správy je základní metodou pro sdílení veřejných informací mezi veřejnoprávními subjekty navzájem i pro sdílení veřejných údajů mezi veřejnoprávní a soukromoprávní sférou v ČR. Veřejný datový fond se od pouhé publikace automatizovaně čitelných otevřených dat posune též k publikaci právně závazných, platných a pravidelně aktualizovaných datových sad s jasně definovanou zodpovědností OVS za takové sady.”

Legislativní ukotvení veřejného datového fondu

VDF není v legislativě jako pojem explicitně zmíněn. Legislativa pouze zavádí pojmy důležité v rámci VDF a pravidla evidence a sdílení údajů ve VDF. VDF je ukotven v zákoně č. 111/2009 Sb., o základních registrech (ZoZR), zákoně č. 106/1999 Sb., o svobodném přístupu k informacím (InfoZ) a zákoně č. 365/2000 Sb. o informačních systémech veřejné správy (ZoISVS).

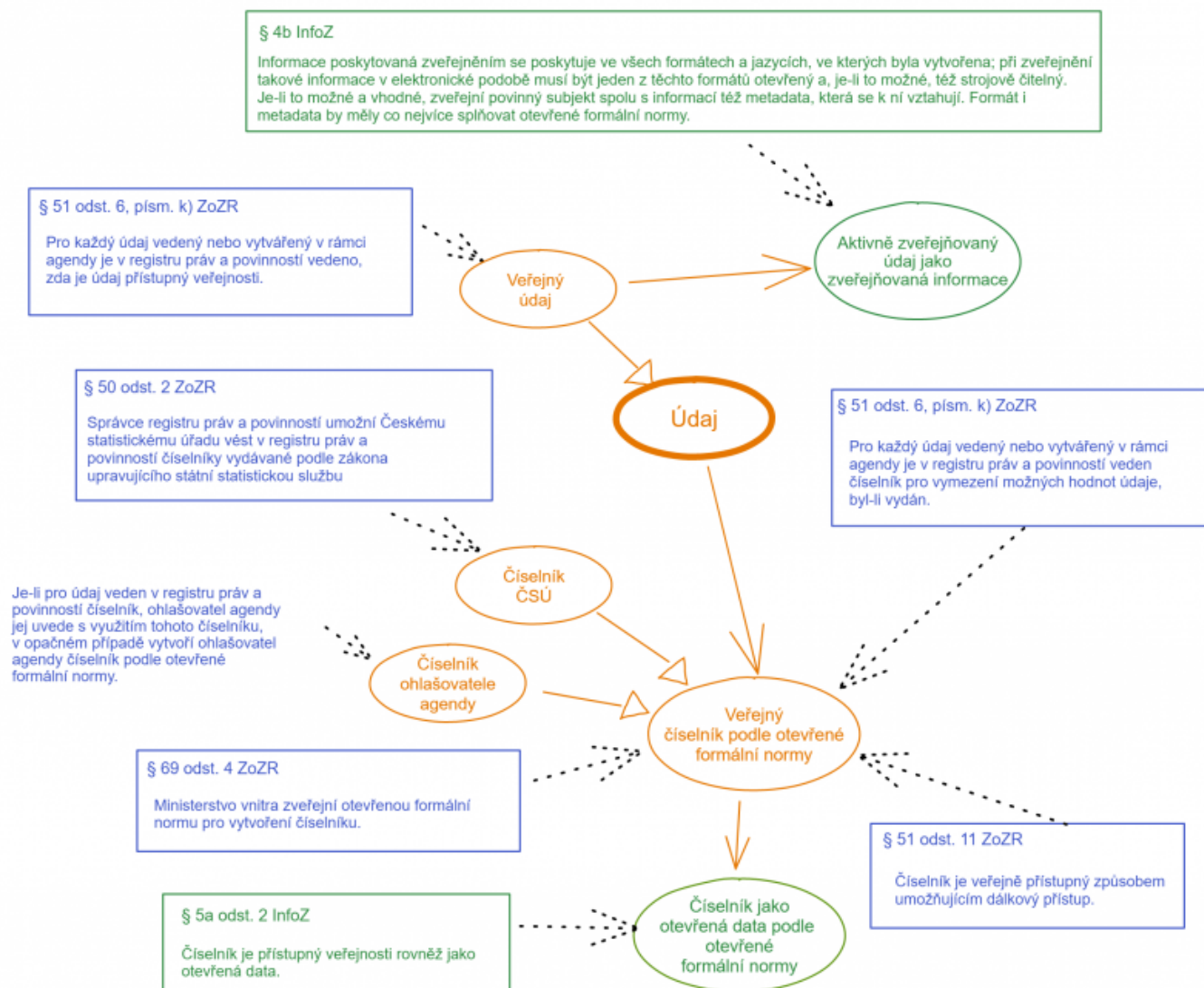
Základním pojmem je údaj vedený nebo vytvářený v rámci agendy. Podle § 51 odst. 6, písm. k) ZoZR je pro každý údaj vedený nebo vytvářený v rámci agendy v RPP vedena jeho přístupnost veřejnosti a v případě, že údaj není přístupný veřejnosti, číslo a název právního předpisu a označení jeho ustanovení (v budoucnu jako odkaz do eSeL), na jehož základě není údaj přístupný veřejnosti (resp. seznam takových referencí na ustanovení právních předpisů). Pokud tedy neexistují legislativní překážky ke zveřejnění údaje, jedná se o údaj přístupný veřejnosti (dále jen veřejný údaj). Veřejný údaj je speciálním případem údaje.

Způsob poskytnutí údaje zveřejněním upravuje § 4b InfoZ, dle kterého se veřejný údaj poskytuje jako informace ve všech formátech a jazycích, ve kterých byla informace vytvořena; při zveřejnění takové informace v elektronické podobě musí být jeden z těchto formátů otevřený a, je-li to možné, též strojově čitelný. Je-li to možné a vhodné, zveřejní povinný subjekt spolu s informací též metadata, která se k ní vztahují. Formát i metadata by měly co nejvíce splňovat otevřené formální normy.

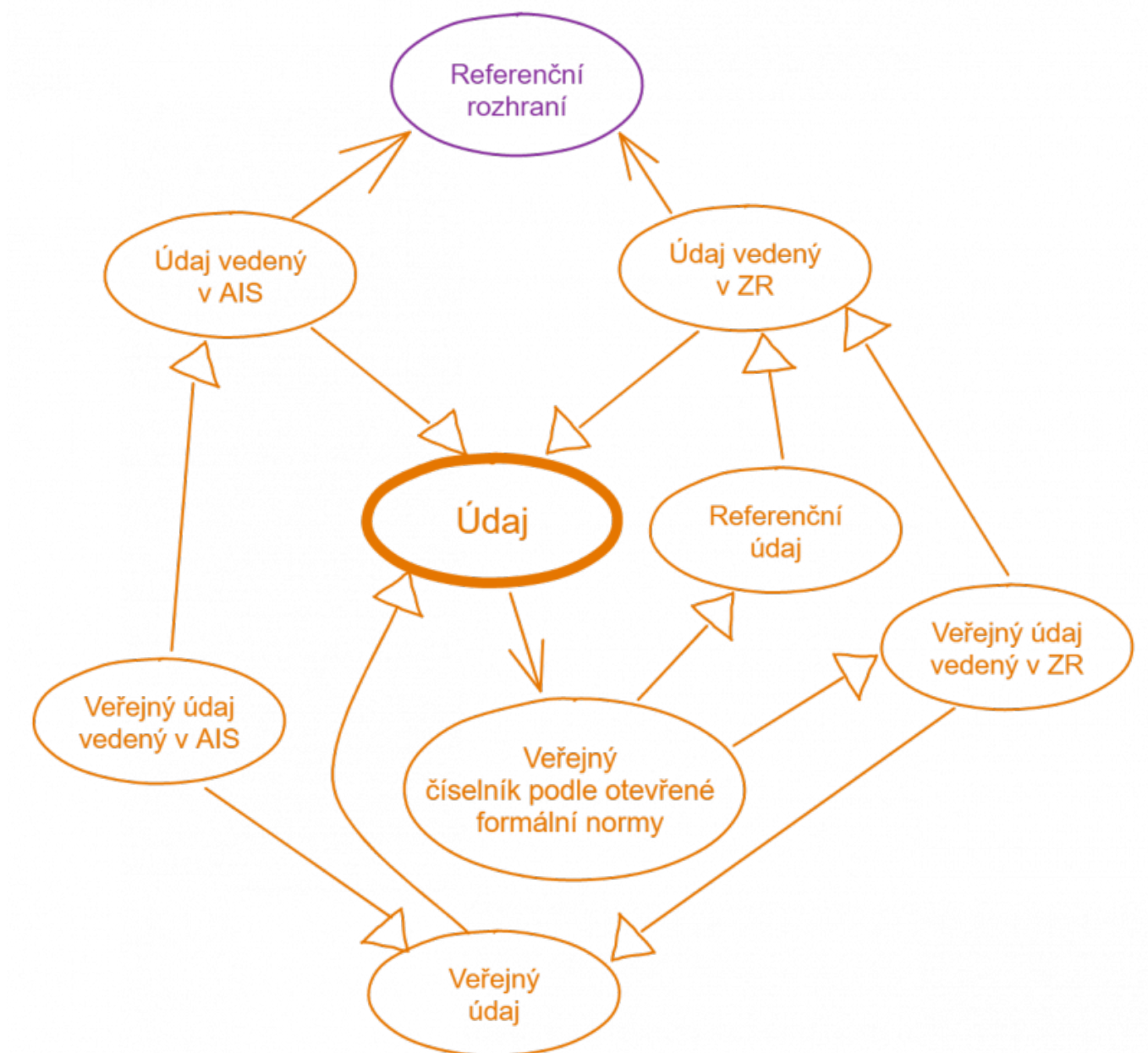
V případě údajů spravovaných v AIS se z principu jedná o informace vedené ve strukturované podobě ve vnitřních databázích informačních systémů, které jsou spravovány databázovými systémy. Databázový systém má rozhraní pro získání obsahu údajů, prostřednictvím kterého lze obsah údajů ve strukturované podobě získat. Jejich zveřejnění v otevřeném a strojově čitelném formátu je tedy technicky možné vždy. Zároveň je možné zveřejnit spolu s informací i metadata a při volbě formátu zveřejněné informace a reprezentaci metadat plně postupovat podle otevřených formálních norem. Z důvodu dosažení interoperability mezi informačními systémy veřejné správy a také z důvodu dosažení přeshraniční interoperability je nutné při zveřejňování údajů postupovat podle [otevřených formálních norem vydávaných MV ČR](#) a při zveřejňování metadat je nutné postupovat dle [otevřené formální normy rozhraní katalogů otevřených dat](#). Jinými slovy, veřejné údaje jsou zveřejňovány v otevřeném a strojově čitelném formátu dle otevřených formálních norem vydávaných MV ČR a jsou katalogizovány v Národním katalogu otevřených dat (NKOD).

Dále je dle § 51 odst. 6, písm. k) ZoZR pro údaj vedený nebo vytvářený v rámci agendy, jehož možné hodnoty jsou vymezeny číselníkem, tento číselník veden v RPP. Podle § 69 odst. 4 ZoZR MV ČR zveřejní otevřenou formální normu pro vytvoření číselníku a podle § 51 odst. 11 ZoZR jsou tyto číselníky veřejně přístupné způsobem umožňujícím dálkový přístup. Číselník pro vymezení možných hodnot údajů pro účely tohoto dokumentu nazýváme veřejný číselník. Ze spojení s § 5a odst. 2 InfoZ pak vyplývá, že veřejný číselník musí být zveřejněn jako otevřená data podle otevřené formální normy vyplývající z § 69 odst. 4 ZoZR.

Z pohledu ZoZR odlišujeme dva možné typy veřejných číselníků. Dle § 50 odst. 2 ZoZR to jsou číselníky ČSÚ. Dle § 54 odst. 1, písm. a) ZoZR to jsou číselníky ohlašovatelů agend. Jinými slovy, možné hodnoty údaje lze vymezit buď číselníkem ČSÚ nebo číselníkem ohlašovatele agendy. Z § 54 odst. 1, písm. a) ZoZR vyplývá, že primárně musí být volen číselník, který již je v RPP veden. Pouze pokud žádný vhodný číselník v RPP veden není, uvádí ohlašovatel agendy svůj vlastní číselník, tj. nový číselník ohlašovatele agendy.



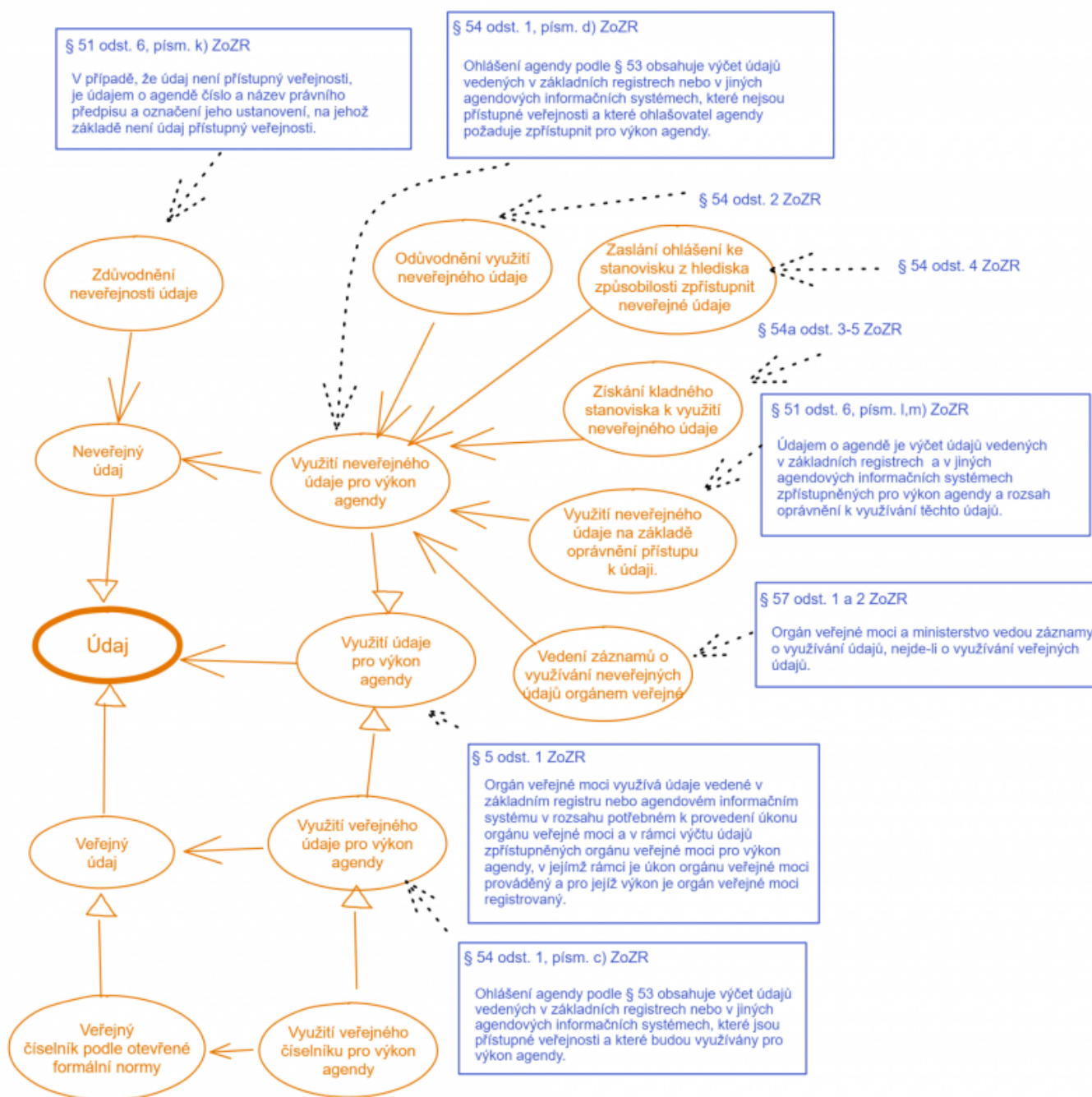
Jak bylo zmíněno na začátku dokumentu, rozlišení údajů na veřejné a neveřejné provádíme z pohledu sdílení údajů. Ať už se jedná o údaje veřejné nebo neveřejné, sdílením se podle § 2 písm. o) ZoISVS rozumí umožnění přístupu k daným datům prostřednictvím referenčního rozhraní.



Podle § 51 odst. 6, písm. k) ZoZR musí ohlašovatel agendy zdůvodnit neveřejnost každého jednotlivého údaje, který není přístupný veřejnosti.

Pro zajištění přístupu k neveřejnému údaji je dle § 54 odst. 1, písm. d) ZoZR nutno projít administrativní procedurou vedoucí k získání kladného stanoviska k využití neveřejného údaje. Je také nutno dle § 51 odst. 6, písm. l,m) ZoZR evidovat údaje zpřístupněné pro výkon agendy a dle § 57 odst. 1 a 2 ZoZR je nutno vést záznamy o využívání neveřejných údajů.

Naproti tomu využívání veřejných údajů je spojeno s nižší administrativní zátěží. Je pouze nutné podle § 54 odst. 1, písm. c) ZoZR při ohlášení agendy uvést, jaké veřejné údaje budou využívány pro výkon agendy.



Principy veřejného datového fondu

Obecným výchozím principem VDF je [princip P13 eGovernmentu “Otevřená data jako standard” \(Open Data by Default\)](#):

“Veřejné údaje evidované orgány veřejné správy ve spravovaných ISVS musí být zveřejňovány jako otevřená data. Pro neveřejné údaje musí být jako otevřená data zveřejňována jejich anonymizovaná podoba, souhrn nebo statistika. V případě, že orgány veřejné správy sdílejí veřejné údaje (včetně anonymizované podoby neveřejných údajů, souhrnů nebo statistik) musí je sdílet jako otevřená data.”

VDF zpřístupňuje veřejné registrované údaje jednotlivým OVM a SPUÚ pro čtení bez omezení přístupu. Navíc jsou všechny údaje přístupné z VDF dostupné ve stejné podobě také prostřednictvím otevřeného přístupu, a to bez výjimky. V obou případech zpřístupnění, tj. prostřednictvím VDF a prostřednictvím otevřeného přístupu, jsou údaje přístupné jako otevřená data dle § 3 odst. 11 InfoZ.

Pro vymezení navazujících pravidel VDF a navržení jeho celkové architektury slouží čtyři základní principy.

- P1 (distribuovanost) - VDF zastřešuje a popisuje skutečné datové zdroje poskytující údaje prostřednictvím referenčního rozhraní a stanovuje pravidla poskytování a čerpání údajů, ale nevynucuje jejich centralizaci na jedno místo.
- P2 (garance) - pro OVM a SPUÚ čerpající údaje z VDF je garantována technická dostupnost a formální správnost (věcná správnost, úplnost, platnost a pravidelná aktualizace) těchto údajů.
- P3 (otevřená data) - údaje přístupné z VDF jsou zpřístupněny jako otevřená data dle § 3 odst. 11 InfoZ bez výjimek.
- P4 (interoperabilita) - údaje jsou z VDF zpřístupněny v podobě, která zajišťuje schopnost různých programových vybavení vzájemně si poskytovat služby a efektivně spolupracovat.

Princip P1 - distribuovanost

VDF zastřešuje skutečné datové zdroje, které nemusí být centralizovány v jednom jediném úložišti. VDF spravuje a organizuje metadata o datových sadách, ve kterých jsou údaje z datových zdrojů zpřístupněny. Mezi tato metadata zejména patří:

- základní charakteristika datové sady (název, popis apod.),
- dokumentace datové sady,
- popis technického zpřístupnění datové sady.

Metadata o datových sadách jsou evidovány v podobě katalogizačních záznamů v NKOD.

Princip P2 - garance

OVM garantuje zpřístupnění veškerých veřejných registrovaných údajů, jejichž je autoritativním zdrojem, prostřednictvím referenčního rozhraní do VDF a na svůj náklad. Údaje dostupné z VDF musí být navíc pro účely čerpání OVM a SPUÚ dostupné, správné, úplné, platné a pravidelně aktualizované s jasně definovanou odpovědností poskytujícího OVM za poskytnutý obsah.

Pro údaje ve VDF platí:

- OVM může do VDF poskytovat pouze údaje, u kterých je autoritativním zdrojem.
- OVM poskytující údaje do VDF garantuje správnost, úplnost, platnost a aktuálnost pro čerpající OVM a SPUÚ.
- Infrastruktura VDF zajišťuje jejich dostupnost.
- Pokud OVM či SPUÚ čerpá údaje z VDF, považuje je za správné, úplné, platné a aktuální a nemusí tyto jejich vlastnosti ověřovat.
- Pokud OVM či SPUÚ využívá údaje dostupné ve VDF, přičemž je nezískal přímo z VDF, ale nějakým jiným způsobem, nemůže tyto údaje považovat za správné, úplné, platné ani aktuální.
- Při změně údajů dostupných ve VDF jsou notifikovány všechny OVM a SPUÚ, které údaje využívají.

Princip P3 - otevřená data

Údaje dostupné ve VDF jsou v totožné podobě také povinně publikovány jako otevřená data dle § 3 odst. 11 InfoZ. VDF tedy chápeme jako prostředek pro přístup jednotlivých OVM a SPUÚ ke garantovaným otevřeným datům.

Princip P4 - interoperabilita

Dva softwarové systémy jsou interoperabilní, pokud jsou schopné si vyměňovat informace a vzájemně si rozumět. V kontextu VDF, jeho širokého využití a velkého počtu informačních systémů je proto nutné zajistit

interoperabilitu na úrovni dat reprezentujících vyměňované údaje. Datovou interoperabilitu lze charakterizovat jako:

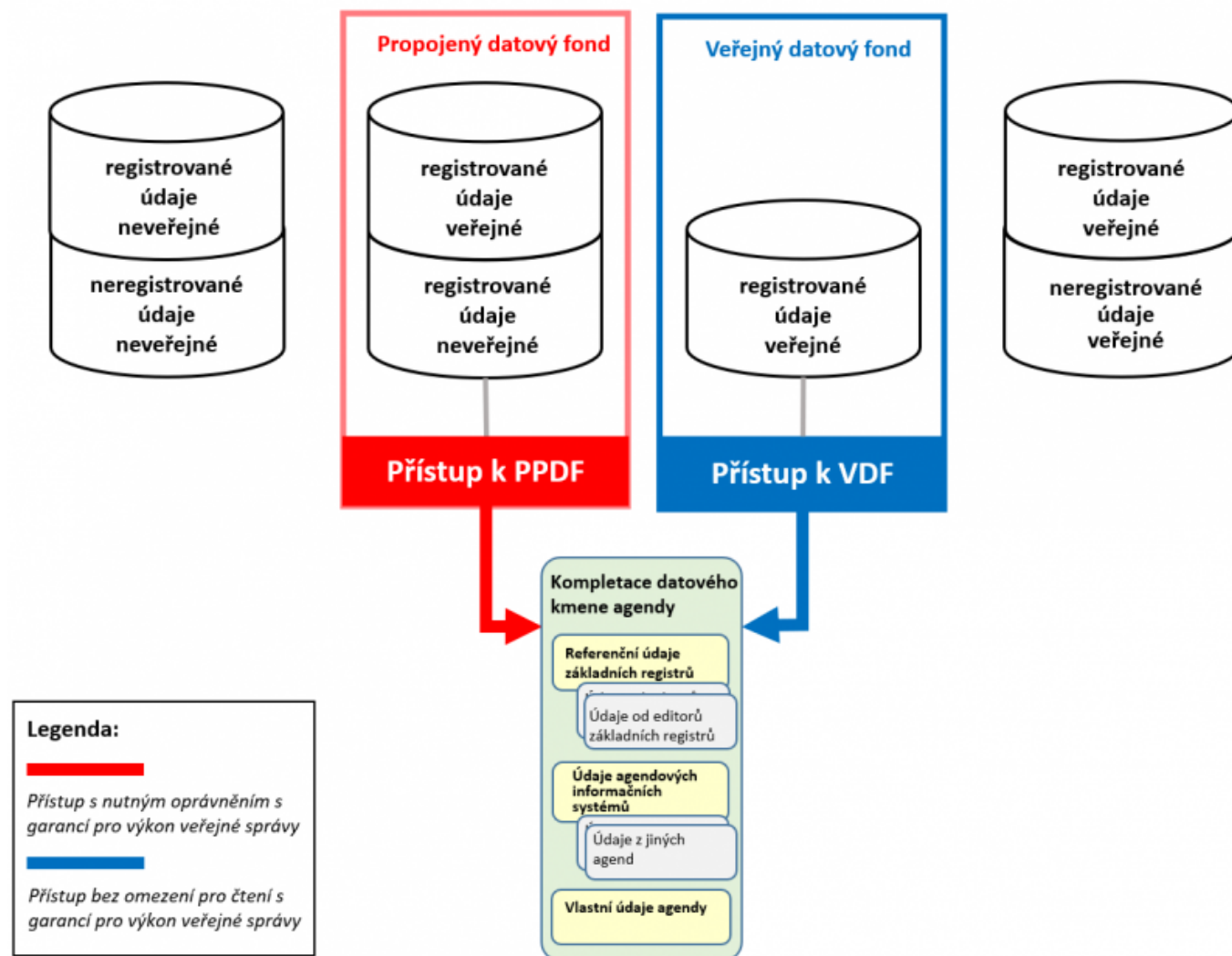
- technickou – použití standardních webových technologií pro perzistentní (tj. trvalou) identifikaci údajů a jejich výměnu,
- syntaktickou – specifikace a dodržování konkrétních formátů dat, komunikačních protokolů a dalších vlastností systémů, které umožňují přistupovat k datům z různých zdrojů standardním způsobem,
- sémantickou – způsob popisu a mapování datových položek, společné slovníky a jednotná klasifikace pojmů.

Ve VDF je technická, syntaktická a sémantická interoperabilita zajištěna prostřednictvím otevřených formálních norem vydávaných MV ČR pro různé typy dat poskytovaných do VDF.

Začlenění veřejného datového fondu do výkonu veřejné správy

Využití veřejného datového fondu pro doplnění datového kmene agendy

Základem pro výkon agendy je její datový kmen. Kompletace datového kmene agendy pro konkrétní výkon je zajišťována prostřednictvím PPDF a VDF. Výměna neveřejných registrovaných údajů je prováděna prostřednictvím PPDF ve vazbě na konkrétní subjekt práva evidovaný v základních registrech na základě oprávnění evidovaných v RPP. Veřejné registrované údaje jsou vyměňovány prostřednictvím VDF. Výměna prostřednictvím VDF je spojena s jednodušším administrativním procesem a oproti výměně prostřednictvím PPDF umožňuje dávkovou výměnu údajů v podobě celého obsahu datových sad s údaji bez vazby na konkrétní subjekt práva. Je-li to nezbytně nutné, lze veřejné registrované údaje sdílet i prostřednictvím PPDF, ale vždy jen na základě oprávnění a ve vazbě na konkrétní subjekt práva.

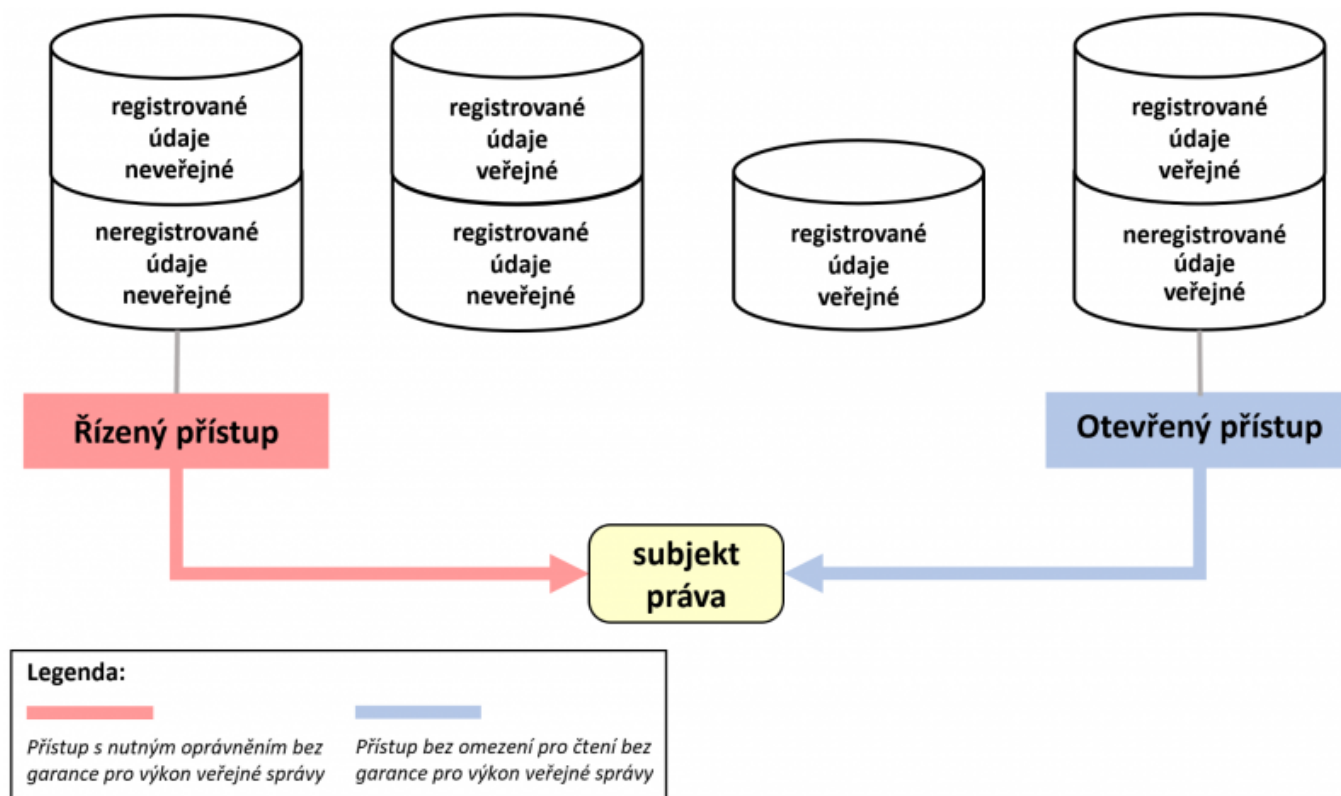


Přístup veřejnosti k údajům

K údajům veřejné správy má možnost přístupu ke čtení i veřejnost.

Přístup veřejnosti k veřejným údajům je zajišťován prostřednictvím otevřeného přístupu. Ten je charakteristický tím, že veřejné údaje jsou pro čtení zpřístupněny anonymním uživatelům z veřejného internetu bez jakéhokoli omezení.

Přístup veřejnosti k neveřejným údajům je realizován prostřednictvím řízeného přístupu. Pro tento přístup ale platí, že je možný pouze na základě definovaných oprávnění konkrétním subjektům práva. Pravidla řízeného přístupu však budou teprve popsána a zde se jim dále nevěnujeme.



Základní struktura veřejného datového fondu

Vnitřní architektura VDF je kromě samotných veřejných registrovaných údajů datového fondu VS ČR tvořena také dílčími nástroji. Dílčí nástroje jsou primárně určeny pro ukládání obsahu datových sad a práci s metadaty datových sad. Dále je tvořen souborem pravidel, postupů a doporučení, které slouží k zabezpečení vyžadovaných vlastností zpřístupněných údajů, k zajištění jejich správné publikace a vytvoření podmínek pro využití publikovaných údajů při výkonu veřejné správy. Architektura VDF je zcela otevřená a připravená pro její další rozšiřování v závislosti na požadavcích na další funkcionality, a také pro podporu sémantické interoperability.

Přístup k údajům prostřednictvím VDF je technicky zajišťován jako

- přístup k datovým souborům s obsahem datových sad ke stažení (povinně)
- přístup ke konkrétní položce datové sady prostřednictvím REST API (volitelně)
- dotazování nad obsahem datové sady prostřednictvím SPARQL API (volitelně)

Přístup je zajištěn pomocí služeb referenčního rozhraní. Pro účely PPDF jsou služby referenčního rozhraní poskytovány prostřednictvím ISZR a ISSS. Oba ale předpokládají přístup k údajům o konkrétním subjektu práva vedeném v ZR a s potřebnými oprávněními. VDF však zavádí dávkový přístup k obsahu jednoho nebo více údajů bez vazby na konkrétní subjekt práva a to ke kompletnímu obsahu nebo k obsahu vázanému na jednu datovou položku. Za tímto účelem musí být referenční rozhraní rozšířeno o nové služby pro:

- čtení publikovaného datového souboru s obsahem datové sady
- čtení publikované položky datové sady
- dotazování nad obsahem datové sady

Tyto nové služby nemohou být z výše uvedených důvodů poskytovány prostřednictvím ISZR ani ISSS a proto je zavedena nová třetí komponenta referenčního rozhraní *Informační systém garantovaných otevřených dat (ISGOD)*. ISGOD zajišťuje jednotlivým OVM a SPUÚ přístup ke garantovaným otevřeným datům ve VDF prostřednictvím referenčního rozhraní.

Do VDF publikují jednotlivé AIS i ZR. Dále tedy v textu hovoříme o publikujícím ISVS, ale vždy je tím myšleno publikující AIS či ZR.

ISGOD je logickým zastřešením několika aplikačních komponent tvořících prostředí VDF. Komponenty jsou popsány v následujících podkapitolách.

Úložiště datových sad

Úložiště datových sad slouží k ukládání obsahu registrovaných veřejných údajů vedených v daném publikujícím ISVS v podobě distribucí datových sad. Pro každý publikující ISVS je vytvořeno úložiště, jehož správcem je správce ISVS. Úložiště datových sad zajišťuje:

- kontrolu souladu nahrávaného obsahu s otevřenými formálními normami
- transformaci obsahu do všech podob daných otevřenými formálními normami s pomocí transformačních skriptů definovaných otevřenými formálními normami
- zpřístupnění těchto podob jako distribuce datových sad, a to jako
 - datové soubory ke stažení (povinně)
 - API umožňující přistoupit ke každé jednotlivé položce datové sady (nepovinně)
 - API umožňující dotazování nad obsahem uložených datových sad pomocí dotazovacího jazyka SPARQL (nepovinně)
- garantovanou dostupnost distribucí datových sad při přístupu prostřednictvím ISGOD
 - požadovaná úroveň dostupnosti distribucí je stejná jako požadovaná úroveň dostupnosti ostatních služeb ISZR a ISSS
- negarantovanou dostupnost distribucí datových sad při přístupu prostřednictvím veřejného internetu

Úložiště datových sad daného publikujícího ISVS není novým ISVS, ale je součástí publikujícího ISVS.

Softwarový nástroj úložiště datových sad může správce ISVS vyvinout vlastní nebo může využít volně dostupný open-source nástroj nabízený a udržovaný MV ČR.

Národní katalog otevřených dat

V NKOD jsou evidovány katalogizační záznamy o všech datových sadách dostupných ve VDF. Za katalogizaci datových sad je zodpovědný publikující ISVS, který nahrává obsah do svého úložiště datových sad jako distribuce obsahu datové sady. Publikující ISVS poskytuje API dle [otevřené formální normy rozhraní katalogů otevřených dat](#), prostřednictvím kterého poskytuje katalogizační záznamy o svých publikovaných datových sadách. API zaregistruje správce ISVS v NKOD. NKOD si poté pravidelně katalogizační záznamy načítá.

Čtenáři údajů z VDF mohou v NKOD vyhledávat datové sady a získávat tak metadata popisující přístup k jejich distribucím prostřednictvím VDF. Metadata také popisují přístup k distribucím prostřednictvím otevřeného přístupu.

Správcem NKOD je MV ČR.

Registr práv a povinností

V RPP je evidována přístupnost registrovaných údajů veřejnosti (tj. veřejnost údajů) a veřejné číselníky, které kódují jejich možné hodnoty. Ohlašovatel agendy je povinen označit přístupnost údaje veřejnosti. Také je povinen uvést, zda je údaj kódován veřejným číselníkem. Obsah veřejných údajů a veřejných číselníků je pak sdílený v podobě distribucí datových sad dostupných jako otevřená data prostřednictvím otevřeného přístupu a prostřednictvím VDF.

Pro potřeby otevřeného přístupu i VDF je pro každý veřejných číselník evidováno IRI datové sady v NKOD, ve které je obsah číselníku přístupný. To samé platí pro veřejný údaj, pouze může být datových sad více.

Čtenáři údajů z VDF vyhledávají v RPP evidenci údajů potřebné údaje. V případě veřejných údajů získávají IRI

datových sad v NKOD, s pomocí kterých mohou z NKOD získat metadata popisující přístup k obsahu veřejných údajů. Podobně v případě údajů, jejichž hodnoty jsou kódovány veřejnými číselníky, získávají IRI datových sad v NKOD, s pomocí kterých mohou z NKOD získat metadata popisující přístup k obsahu veřejných číselníků.

V RPP je také pro každou agendu evidován výčet veřejných registrovaných údajů, které jsou využívány pro výkon agendy.

Správcem RPP je MV ČR.

Katalog uživatelů dat

Katalog uživatelů dat eviduje, jaké datové sady z VDF čerpají konkrétní OVM a SPUÚ. Registraci čerpání datové sady provádí OVM či SPUÚ za účelem získávání notifikací o změnách v datové sadě. Registrace je předvyplněna z evidence čtení registrovaných veřejných údajů v RPP.

Registrace mj. obsahuje:

- IRI OVM či SPUÚ
- IRI datové sady
- požadovanou maximální frekvenci notifikací (ihned, hodinová, denní, týdenní, ...)
- požadovaný způsob a technická specifikace notifikace (datová schránka, WebSub protokol)

Správcem katalogu uživatelů dat je MV ČR.

Notifikační hub

Notifikační hub je nástroj zajišťující s pomocí katalogu uživatelů dat notifikační službu, která informuje subjekty registrované v katalogu uživatelů dat ke čtení datových sad z VDF o změnách v datových sadách. Jedná se o automatickou notifikaci o změnách zjištěných při ohlášení změny ze strany poskytovatele ISVS.

Notifikační hub je implementován na bázi mezinárodního standardu [W3C Recommendation WebSub](#).

Správcem Notifikačního hubu je MV ČR.

Směrovací služba

Distribuce datové sady je v otevřených datech zpřístupněna jako datový soubor ke stažení nebo jako API, které umožňuje získat data každé jednotlivé položky datové sady (entity), o níž jsou v datové sadě reprezentovány údaje. Přístup k údajům je realizován s pomocí tzv. dereference identifikátoru položky. V otevřených datech jsou jako identifikátory použity tzv. IRI (Internationalized resource identifier, více viz [Otevřená formální norma pro propojená data](#)). Dereference identifikátoru v podobě IRI znamená přistoupení k IRI prostřednictvím HTTP protokolu podobně jako přistupujeme k webovým stránkám. Výsledkem dereference je strojově nebo lidsky čitelná reprezentace údajů o identifikované položce (v závislosti na HTTP content negotiation).

IRI položky vyplývá z úložiště datových sad, na kterém je datová sada s položkou fyzicky uložena a z [pravidel pro tvorbu IRI](#). To ale činí IRI závislé na konkrétním fyzickém umístění, které se může měnit. Např. může změnit své fyzické umístění celé úložiště tím, že se změní jeho provozovatel nebo je přesunuto do jiného prostředí. Nebo může dojít k rozhodnutí přesunout obsah datové sady do jiného úložiště. Pokud k takové změně dojde, může dojít i ke změně IRI položek. Přes tato IRI ale již mohou být propojeny údaje o prvku či souvisejících prvcích z jiných datových sad ve VDF i v otevřených datech či údajích dostupných přes řízený přístup. Pokud se IRI takto mění, znamená to, že nejsou tzv. perzistentní, tj. trvalé. Pro zajištění perzistentních, tj. trvalých, identifikátorů slouží směrovací a indexační služba.

Pro veřejný údaj je nutné, aby správce ISVS, ze kterého se obsah údaje publikuje, zavedl identifikátory položek datové sady do směrovací služby a určil k nim tzv. referenční identifikátory na doméně gov.cz. Pro položku tak existují dva identifikátory. První je původní identifikátor položky určený správcem ISVS. Vyplyvá z úložiště datových sad, ve kterém je obsah datové sady uložen. Jedná se o IRI v doméně úložiště, při jehož dereferenci úložiště poskytne obsah o identifikované položce. Nazýváme jej lokální identifikátor položky nebo zkráceně *lokální IRI položky*. Druhý identifikátor je nový identifikátor, který je nezávislý na úložišti. Nazýváme jej referenční identifikátor položky ve VDF nebo zkráceně *referenční IRI položky*. Referenční IRI používají pro odkazování se na položku všechny datové sady. Dereference referenčního IRI vede na směrovací službu, která provede přesměrování na lokální IRI položky.

Při změně úložiště datových sad nebo obecně při změně lokálních IRI je upraveno směrování ve směrovací službě. Tím je zajištěna perzistence IRI, která umožňuje trvalé a neměnné propojení údajů o stejné položce napříč různými datovými zdroji nezávisle na poskytovatelích těchto údajů.

Správcem směrovací a indexační služby je MV ČR.

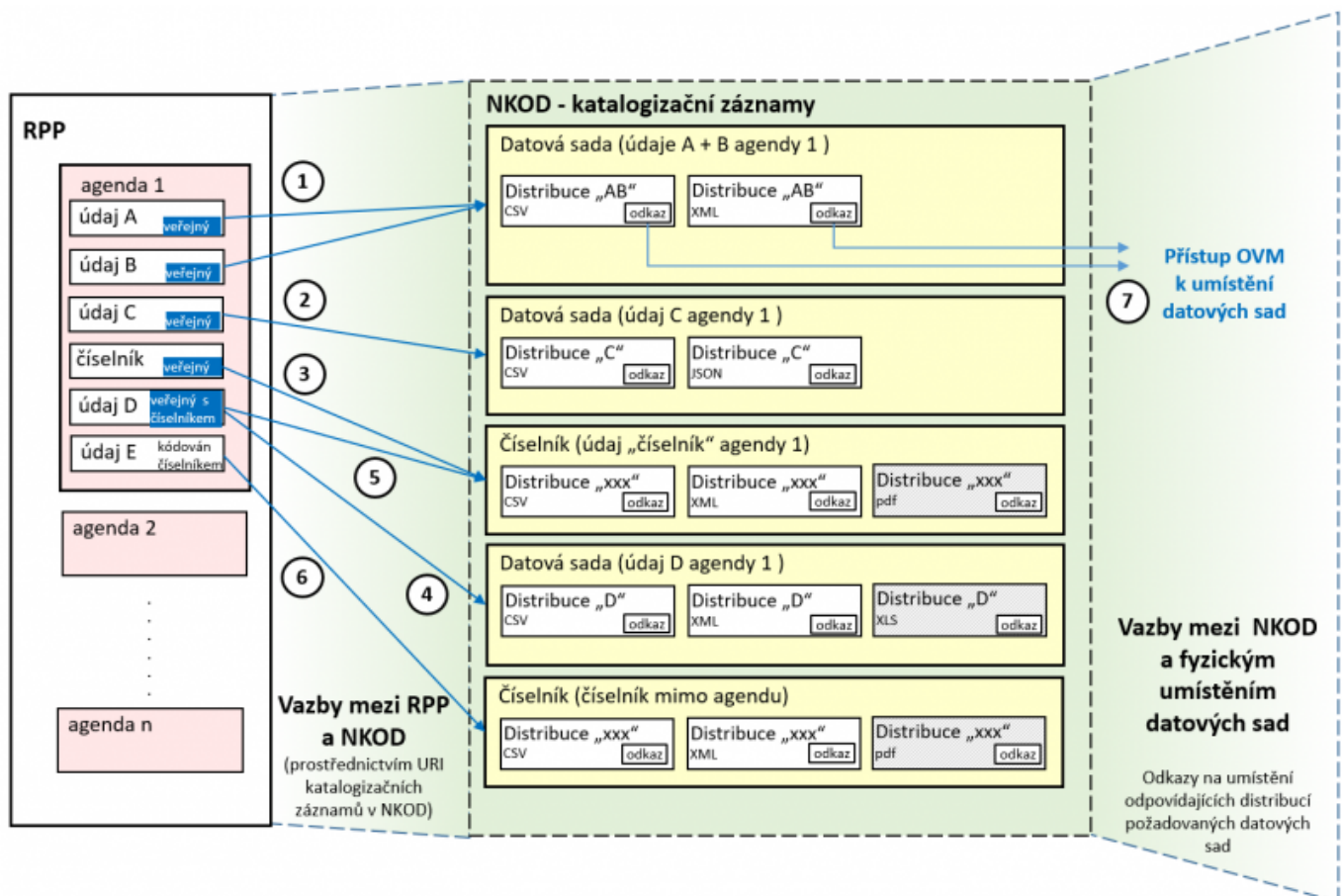
Pravidla veřejného datového fondu

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.

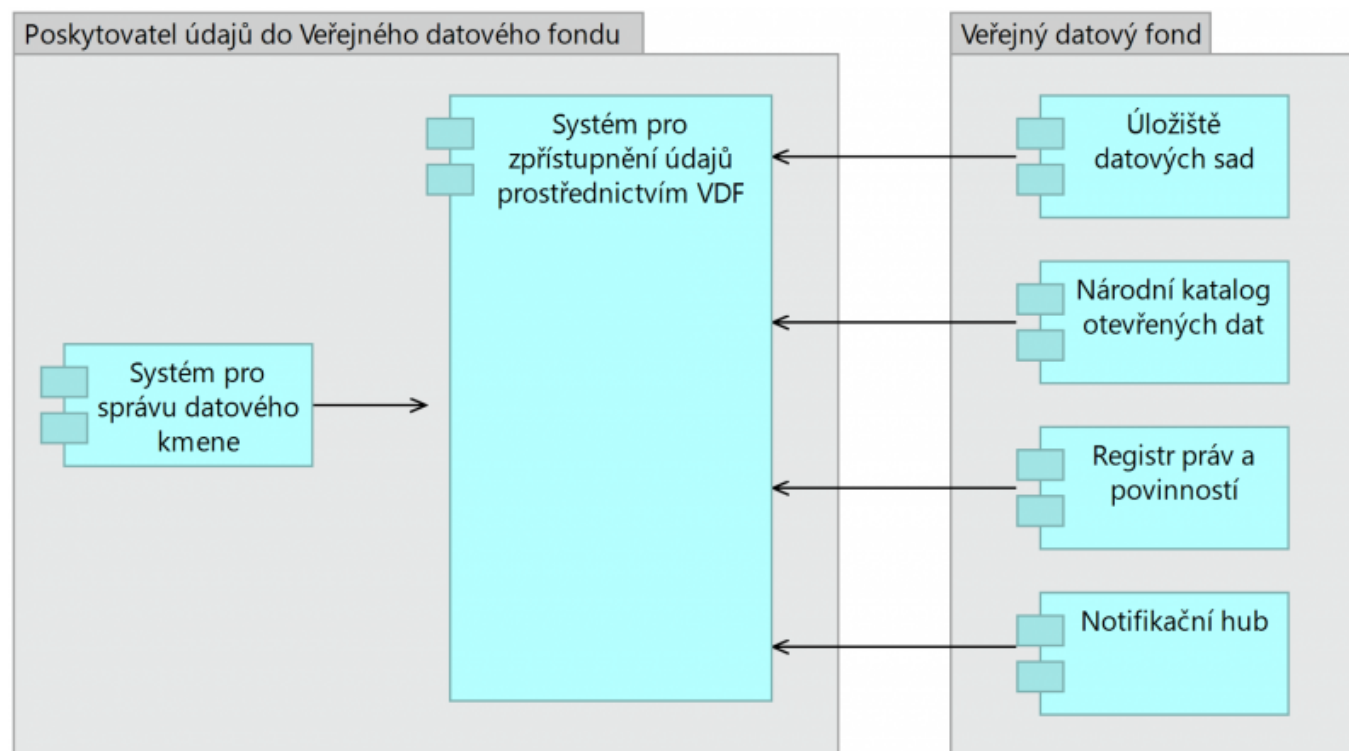


- V RPP je u agendy 1 evidováno, že údaje A i B jsou veřejné. To znamená, že jsou dostupné jako otevřená data prostřednictvím VDF a otevřeného přístupu.
 - Oba údaje jsou dostupné prostřednictvím stejné datové sady.
 - V RPP je v evidenci těchto údajů uvedeno IRI datové sady v NKOD (datová sada se jmenuje „Údaje A + B agendy 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 agendy 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 agendy 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 agendy 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.



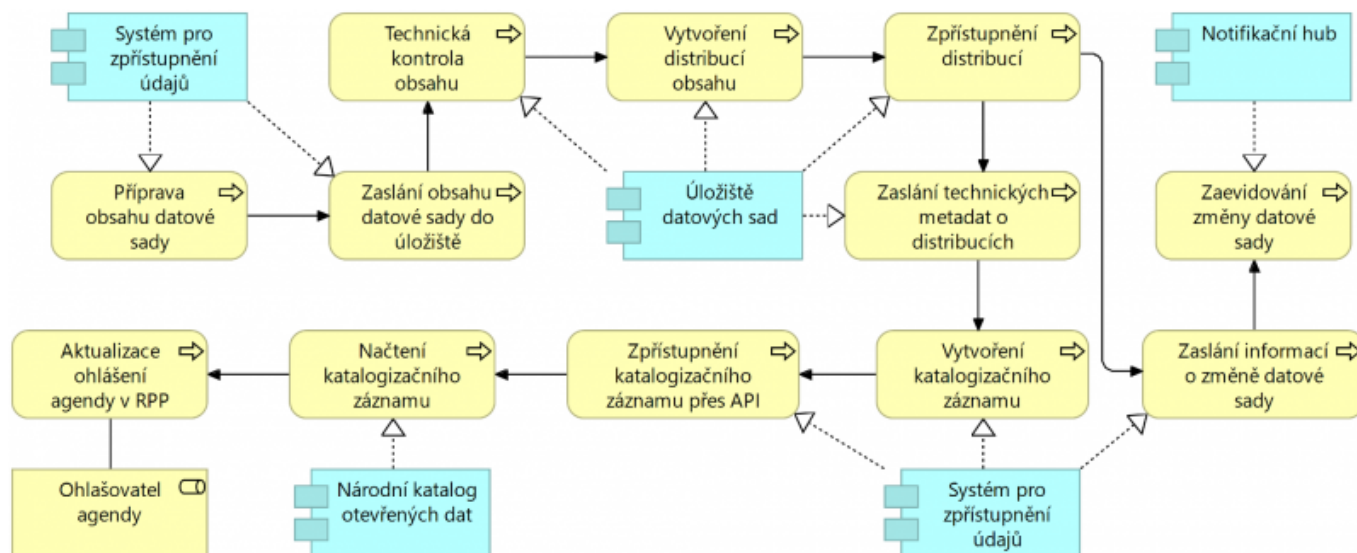
Poskytovatelem údajů do VDF je správce ISVS, ve kterém jsou vedeny registrované veřejné údaje. Tento ISVS je vyznačen na levé straně obrázku jako *systém pro správu datového kmene*, kterým OVM spravuje svůj datový kmen. V praxi se samozřejmě může jednat o více ISVS, zde si pro jednoduchost zobrazujeme 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.
 - Zašle 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 schématům definovaných otevřenými formálními normami (např. vůči JSON nebo XML schématům)
 - V případě špatné syntaxe zašle 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 zaslaného katalogizačního záznamu zaeviduje (zpravidla do 1 dne).
- Notifikační hub
 - Zaeviduje informaci o změně datové sady zaslanou úložištěm datových sad.

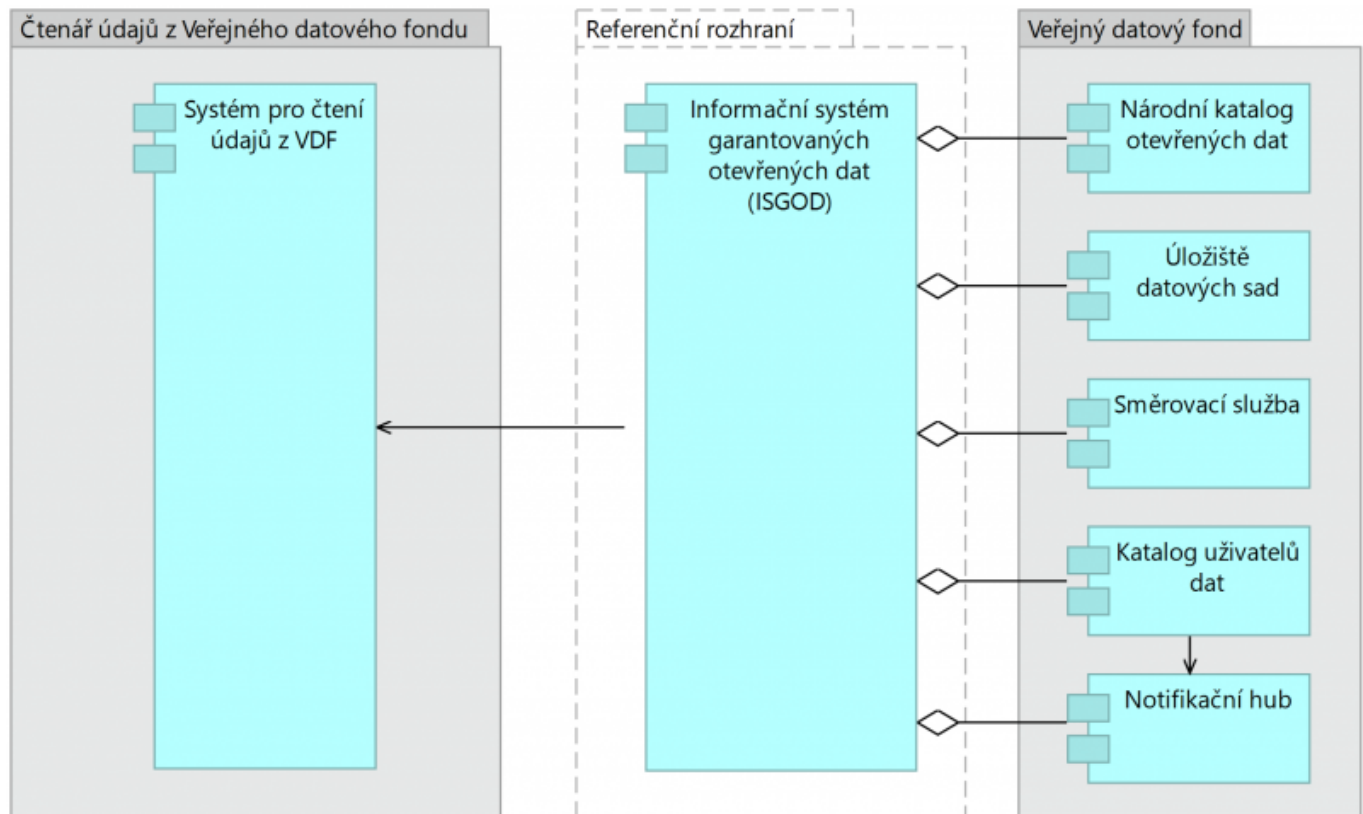
Systém pro zpřístupnění údajů prostřednictvím VDF souvisí s existujícím systémem pro zpřístupnění údajů prostřednictvím PPDF, který zajišťuje poskytování údajů o konkrétním subjektu práva, na který přistupuje čtenářský AIS prostřednictvím PPDF. Systém pro zpřístupnění údajů prostřednictvím VDF (dále 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.

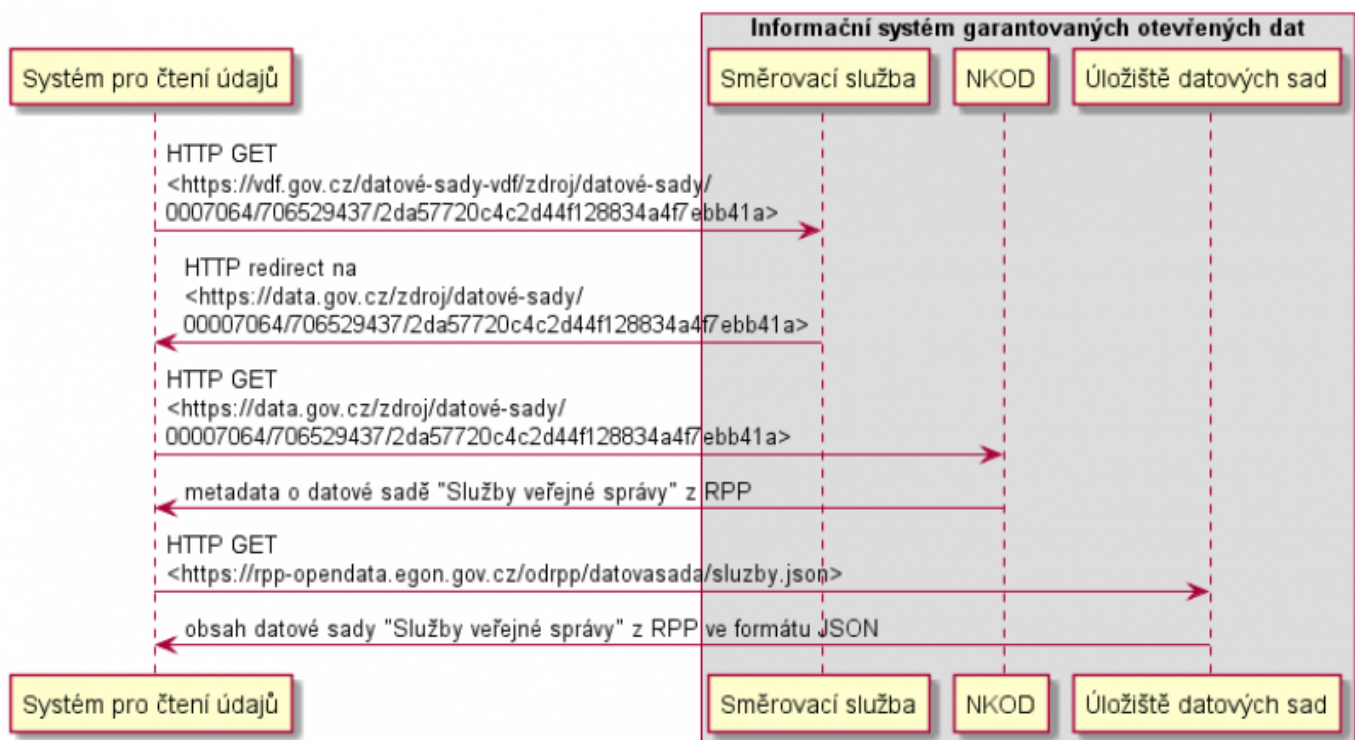
Čtení obsahu datové sady v podobě datového souboru

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 údajem 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:

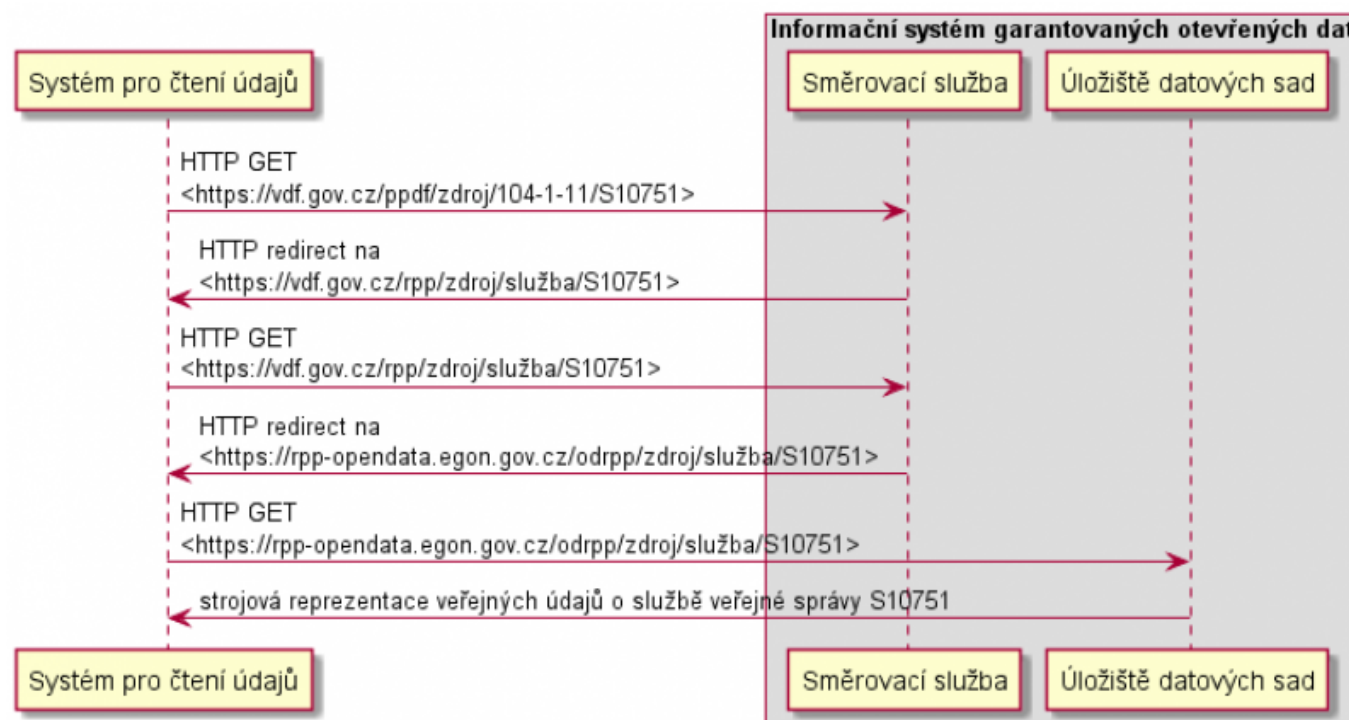
- V předchozích krocích byl přečten z VDF údaj s referenčním IRI jako hodnotou.
- V předchozích krocích byl přečten z PPDF údaj s proprietárním identifikátorem entity jako hodnotou (tj. identifikátor v podobě řetězce, který identifikuje entitu lokálně v rámci ISVS). Referenční IRI položky s veřejnými 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 verze 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>>

Příloha 1: Metodika poskytování a čerpání údajů prostřednictvím VDF

1. Postupy zpřístupňování údajů prostřednictvím VDF
 - Určení údajů poskytovaných do VDF
 - Návrh podoby datových sad a jejich distribucí
 - Příprava informačního systému pro export dat
 - Export údajů do distribucí datových sad
 - Publikace distribucí datových sad
 - Aktualizace údajů, archivace a notifikace
 - Katalogizace datové sady
 - Registrace lokálního katalogu
 - Evidence odkazů na datové sady v RPP
2. Postupy čerpání údajů prostřednictvím VDF
 - Vyhledávání datových sad s údaji
 - Import údajů z datových sad do informačního systému
 - Příjem notifikací o změnách v datových sadách

Přístupy se liší nejen v tom, jaké údaje zpřístupňují a komu je zpřístupňují, ale také v několika dalších aspektech shrnutých v následující tabulce.

Příloha 2: Srovnání vlastností jednotlivých způsobů přístupu

	Přístup k PPDF	Řízený přístup	Přístup k VDF	Otevřený přístup
úroveň garance kvality obsahu	správnost, úplnost, platnost a aktuálnost	správnost, úplnost, platnost a aktuálnost	správnost, úplnost, platnost a aktuálnost	správnost, úplnost, pravidelné aktualizace pro zajištění maximální možné platnosti a aktuálnosti
úroveň garance zpřístupněných údajů pro výkon agend veřejné správy	vysoká dostupnost, formální správnost	bez garance pro výkon agend	vysoká dostupnost, formální správnost	bez garance pro výkon agend

sjednocený popis zpřístupněných údajů	RPP	bez popisu ¹⁾	RPP	bez popisu
centrální popis způsobu zpřístupnění údajů	TSÚA	NKOD	NKOD	NKOD
podoba definice datových formátů pro zpřístupnění	TSÚA	OFN	OFN	OFN
systém pro správu oprávnění	RPP	KUD	přístup bez omezení	přístup bez omezení
systém pro vedení evidence přístupu k údajům	RPP	KUD	KUD	bez evidence
identifikace subjektů a objektů práva	IČ a AIFO pro subjekty, proprietární id pro objekty	IRI	IRI	IRI

1)

Pro definici oprávnění a řízení přístupu je popis nutný. V současné době ale žádný systematický nástroj popisu neexistuje a bude nutné jej vytvořit. Tímto nástrojem bude sémantický slovník pojmů vyjádřený v podobě ontologických modelů právních předpisů a agend. Ten bude navíc sloužit jako zastřešující model a popis všech dat v datových fondech OVM.

From:

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

Permanent link:

<https://archi.gov.cz/playgroud:vdf?rev=1630910837>

Last update: 2021/09/06 08:47

