

DIGITÁLNÍ A INFORMAČNÍ AGENTURA_

Export z Národní architektury eGovernmentu ČR

Obsah

Popis Veřejného datového fondu	3
<i>Legislativní ukotvení veřejného datového fondu</i>	3
<i>Principy veřejného datového fondu</i>	6
<i>Začlenění veřejného datového fondu do výkonu veřejné správy</i>	8
<i>Základní struktura veřejného datového fondu</i>	10

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.

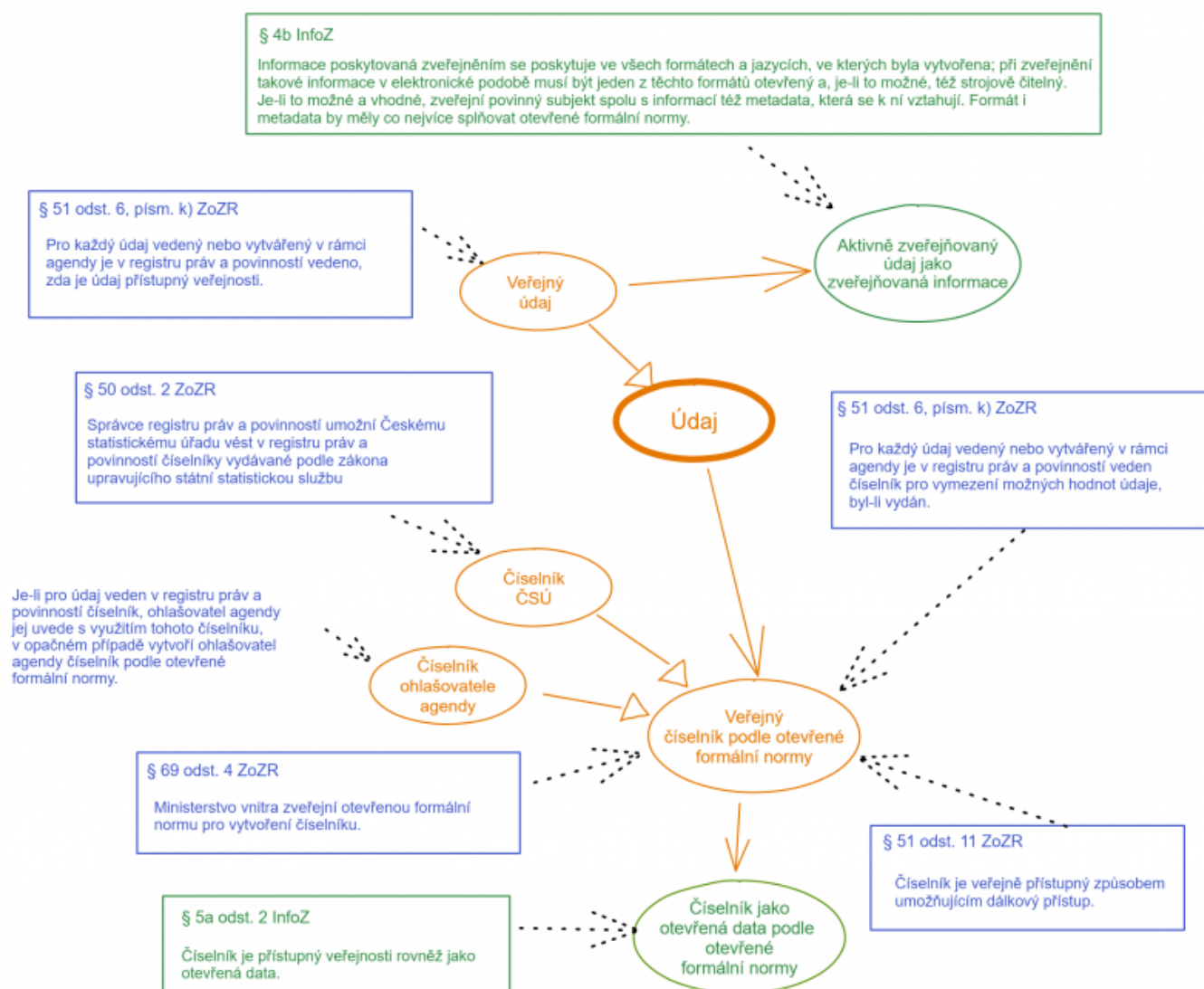
Způsob poskytnutí údaje zveřejněním upravuje § 4b [InfoZ](#), dle kterého se veřejný údaj poskytuje jako informace ve všech formátech a jazycích, ve kterých byla informace vytvořena; při zveřejnění takové informace v elektronické podobě musí být jeden z těchto formátů otevřený a, je-li to možné, též 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).

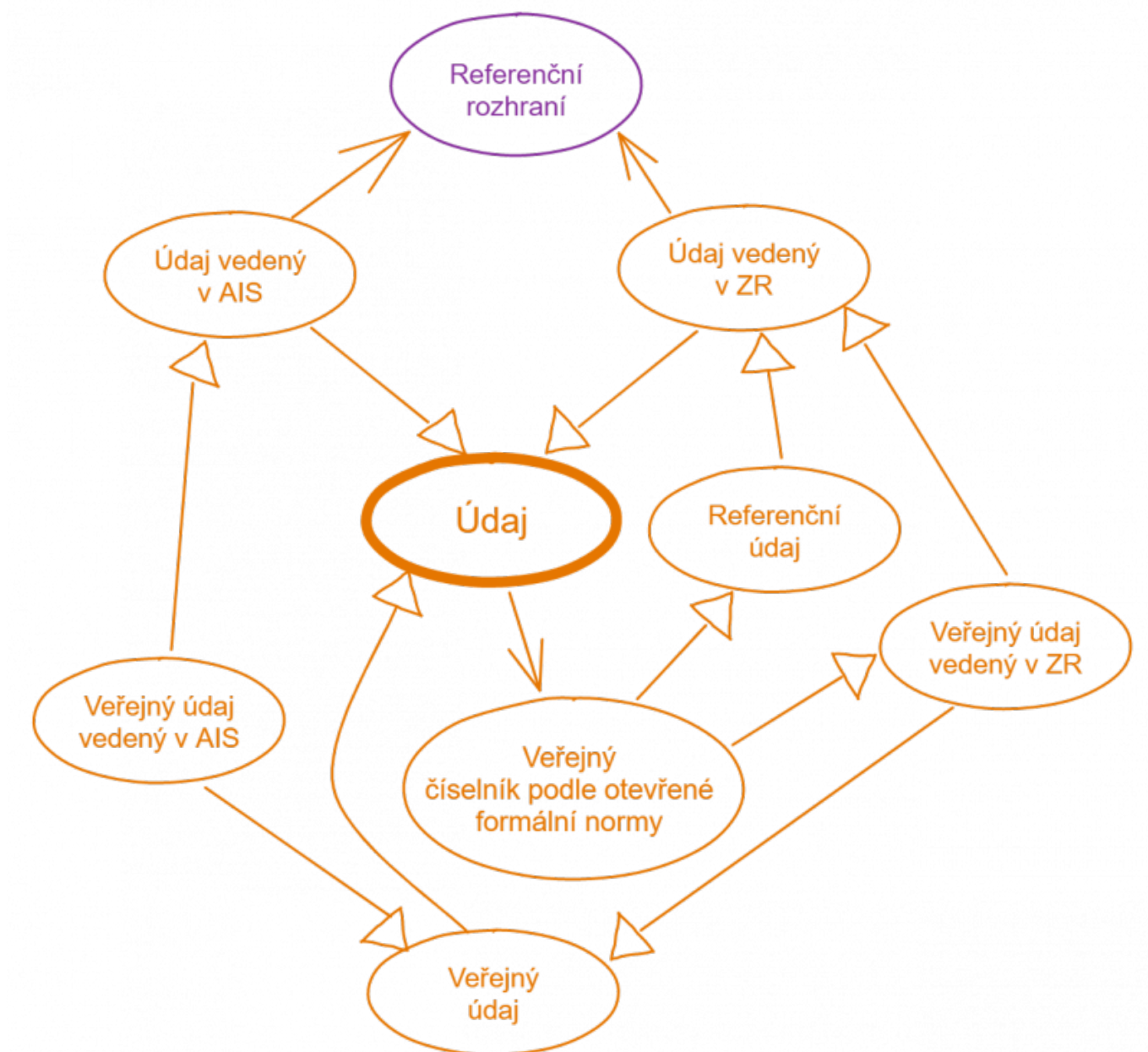
Dále je dle § 51 odst. 6, písm. k) [ZoZR](#) pro údaj vedený nebo vytvářený v rámci agendy, jehož možné hodnoty jsou vymezeny číselníkem, tento číselník veden v [RPP](#). Podle § 69 odst. 4 [ZoZR](#) MV ČR zveřejní otevřenou formální normu pro vytvoření číselníku a podle § 51 odst. 11 [ZoZR](#) jsou 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.



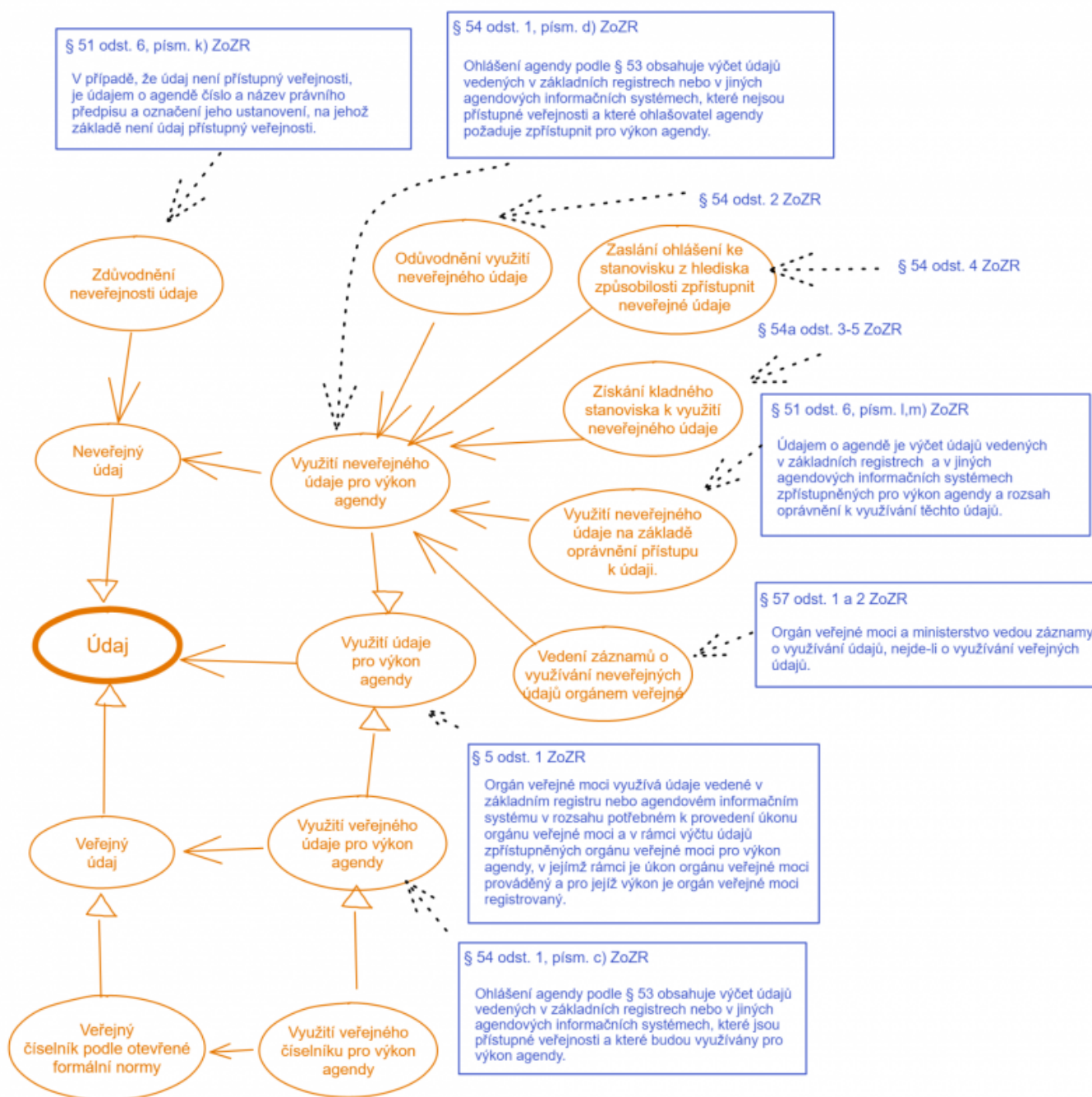
Jak bylo zmíněno na začátku dokumentu, rozlišení údajů na veřejné a neveřejné provádíme z pohledu sdílení údajů. Ať už se jedná o údaje veřejné nebo neveřejné, sdílením se podle § 2 písm. o) **ZoISVS** rozumí umožnění přístupu k daným datům prostřednictvím **referenčního rozhraní**.



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 výjimky.
- 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:

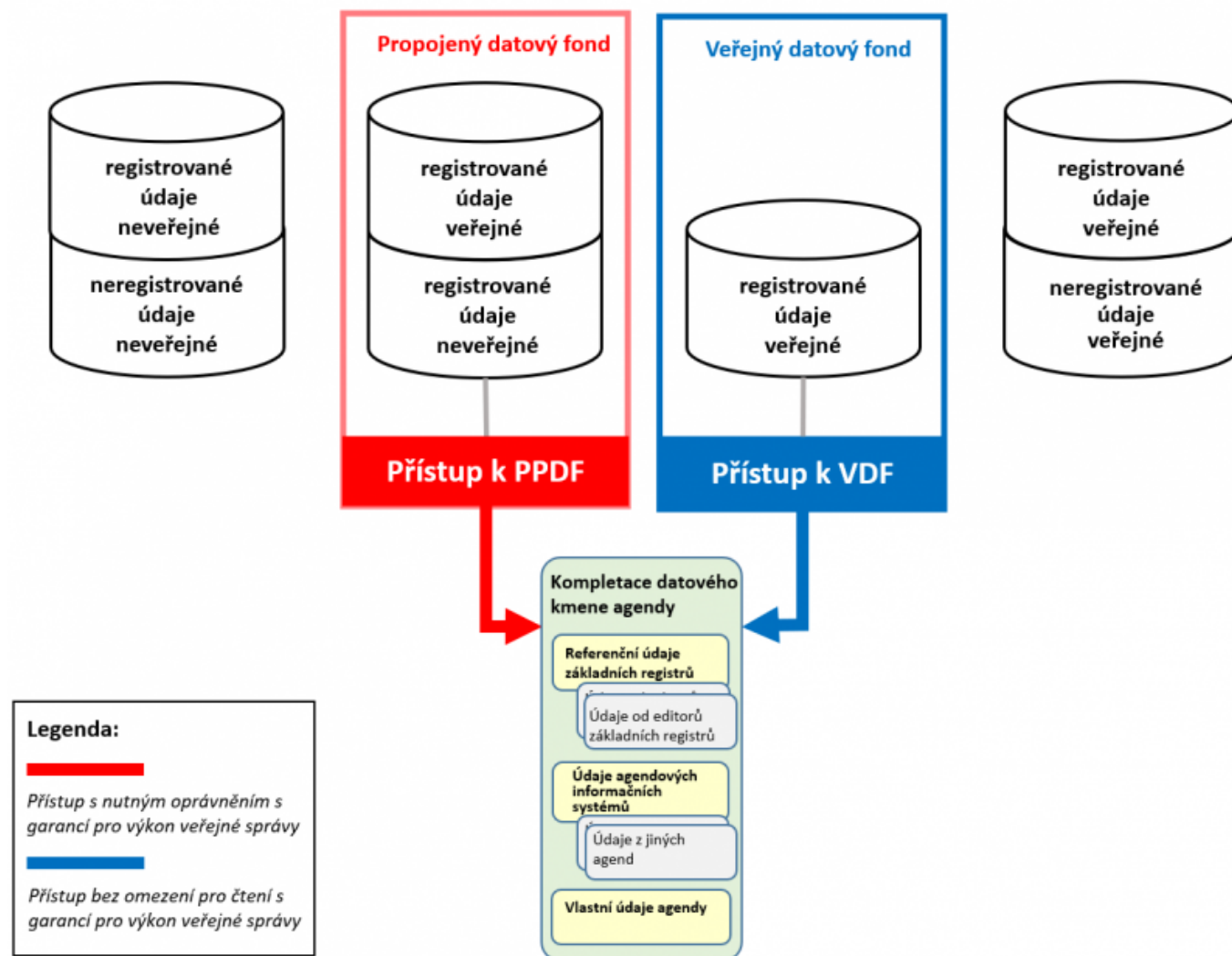
- technickou – použití standardních webových technologií pro perzistentní (tj. trvalou) identifikaci údajů a jejich výměnu,
- syntaktickou – specifikace a dodržování konkrétních formátů dat, komunikačních protokolů a dalších vlastností systémů, které umožňují přistupovat k datům z různých zdrojů standardním 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.

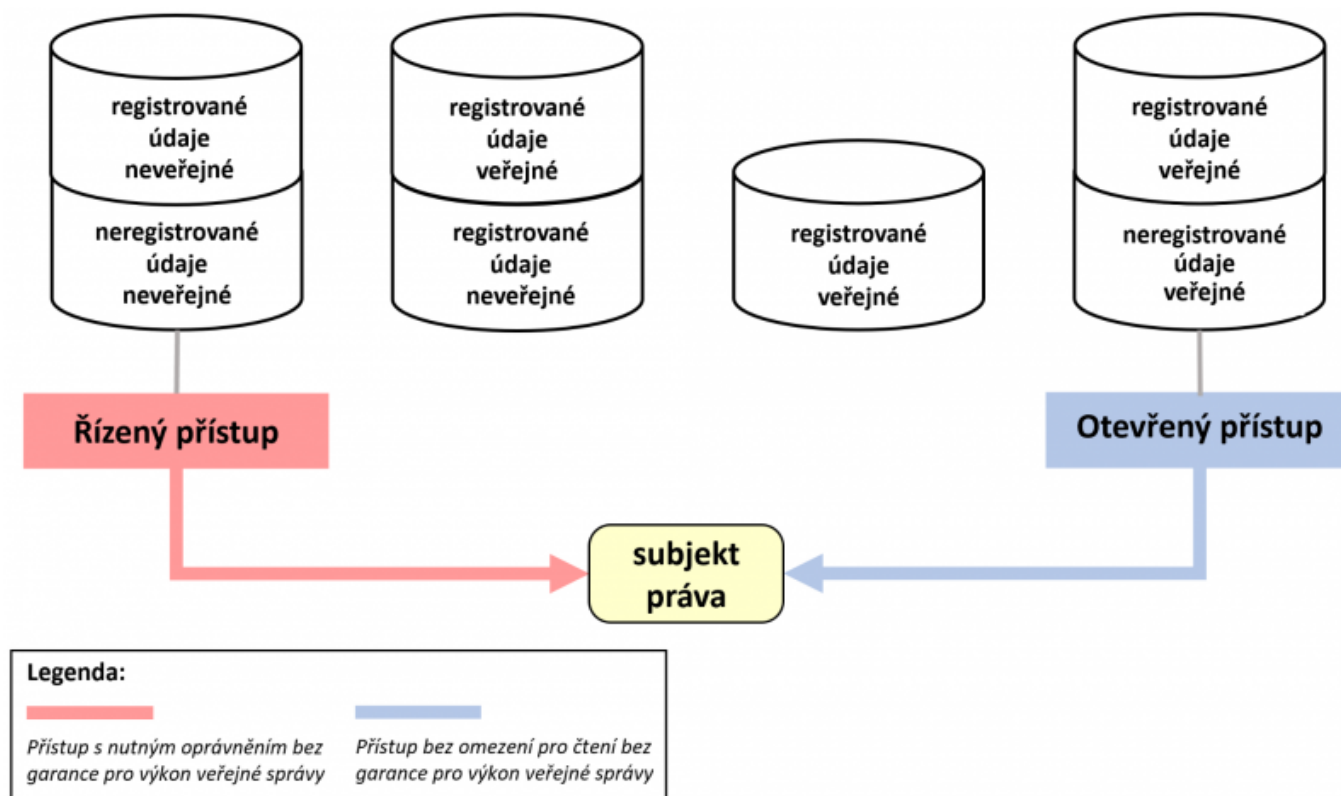


Přístup veřejnosti k údajům

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 jakéhokoliv 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 veřejný 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 poskytujícího 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.

From:

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

Permanent link:

https://archi.gov.cz/nap:popisy:popis_verejny_datovy_fond

Last update: **2021/04/30 10:46**

