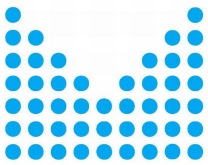
Materiál Ministerstva vnitra



MINISTERSTVO VNITRA
ČESKÉ REPUBLIKY

Export z Národní architektury eGovernmentu ČR

Obsah

<title>Veřejný datový fond</title>



Popis architektury úřadu a veřejné správy ČR po jednotlivých vrstvách architektury a zapracování požadavků do informační koncepce a architektury úřadu je popsán v části Architektura úřadu v kontextu veřejné správy a jejích vrstvách architektury.

Pravidla pro jednotlivé sdílené služby, funkční celky a tematické oblasti jsou popsány v části Způsoby využívání sdílených služeb, funkčních celků a tematických oblastí jednotlivými úřady

Veřejný datový fond (také jako "VDF") je princip vytváření a dotváření obrazu propojeného datového fondu dle principů otevřených dat za účelem podpory sdílení údajů OVS při výkonu veřejné správy i mimo jejich rozsah práv a povinností zachycených v RPP a mimo PPDF.

VDF zastřešuje a reprezentuje skutečné fyzické datové zdroje otevřených dat roztroušené ve webovém prostoru VS. Vlastní zastřešení a vymezení je realizováno nástroji určenými pro správu metadat otevřených dat, což nejen konkretizuje samotný pojem VDF, ale nabízí i prostředky pro další kategorizace v rámci celého datového fondu ČR.

Význam a role VDF se projevuje ve dvou oblastech VS.

První se týká interního režimu VS, kde je hlavním prostředkem sdílení veřejných údajů mezi veřejnoprávními subjekty navzájem a také rozšiřuje vzájemné sdílení informací o ty údaje, které v PPDF vůbec nejsou. Význam VDF v této oblasti nejlépe vystihuje znění dílčího cíle 5.10 z Informační koncepce 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 Open Dat posune též k publikaci právně závazných, platných a pravidelně aktualizovaných datových sad s jasně definovanou zodpovědností OVS za takové sady.”

Druhou důležitou oblastí, ve které VDF sehrává významnou roli, je oblast směřování ČR k otevřenému vládnutí a postupnému zpřístupňování informací VS veřejnosti. VDF je tvořen publikovanými informacemi VS v podobě otevřených dat a tak se stává součástí všech veřejně přístupných otevřených dat v ČR.

Datové zdroje VS ČR

Datový zdroj je způsob přístupu k datům. Data VS ČR jsou přístupná prostřednictvím datových zdrojů v následujících kategoriích:

PPDF - propojený datový fond je popsán zde.

OD - otevřená data - je tvořen všemi publikovanými daty VS ve formátu otevřených dat, která musí splňovat požadavky na otevřená data odpovídající definici uvedené v § 3 odst. 11 zákona č.106/1999 Sb. Součástí otevřených dat jsou také všechna otevřená data samospráv, otevřená data využívaná v konceptu SMART Cities a data publikovaná v rámci VDF.

Pro data ve formátu otevřených dat platí, že jsou:

- Volně přístupná na webu jako datové soubory ke stažení ve strojově čitelném a otevřeném formátu - CSV, XML, JSON, RDF a v dalších formátech s otevřenou specifikací.

- Opatřená podmínkami užití neomezujícími jejich užití.
- Evidovaná v Národním katalogu otevřených dat (NKOD) jako datové sady opatřené přímými odkazy na datové soubory, které je tvoří.
- Opatřená úplnou, volně dostupnou a otevřenou dokumentací.
- Opatřená kontaktem na kurátora pro zpětnou vazbu (chyby, žádost o rozšíření, apod.)
- Jsou publikovány dle otevřených formálních norem ve smyslu § 4b odst. 1 zákona č. 106/1999 Sb. o svobodném přístupu k informacím.

VDF - veřejný datový fond - je podmnožinou otevřených dat. Je tvořen datovými zdroji, které veřejně zpřístupňují data dostupná v PPDF a datovými zdroji, která veřejně zpřístupňují data, jež PPDF vůbec neobsahuje, ve formátu otevřených dat. Jedná se např. o data nutná pro správnou interpretaci dat z PPDF (např. číselníky) nebo o data obsažená v agendách, které nemají ze zákona povinnost publikovat svoje údaje jiným agendám, a tedy ani nejsou součástí PPDF. VDF tedy slouží ke zpřístupnění dat PPDF jednotlivým OVS v těch případech, kdy příslušná OVS nemají možnost tato data sdílet prostřednictvím PPDF (nemají přidělena práva v RPP), nebo takových dat, které vůbec nejsou součástí PPDF a v obou případech taková data ke své práci potřebují. Veškerá publikovaná data ve VDF musí být ve formátu otevřených dat a musí splňovat požadavky kladené na otevřená data dle zákona č.106/1999 Sb. Nad rámec těchto pravidel musí poskytovatel u svých dat publikovaných ve VDF garantovat jejich aktuálnost, dostupnost a správnost tak, aby je ostatní OVS mohly využívat při své práci.

Součástí VDF jsou rovněž data s řízeným přístupem. Jedná se o datové sady, které naplňují všechny stanovené podmínky pro otevřená data s výjimkou požadavku, aby jejich poskytovatel nijak neomezoval jejich následné využití. Příkladem takové praxe je podmínění poskytnutí dat předchozí registrací, aby mohl poskytovatel sledovat, jak je s daty nakládáno. Jedná se o takové případy, kdy úplnému otevření dat brání objektivní důvody vyplývající například z legislativy nebo věcně technického charakteru dané datové sady.

Ačkoli data s řízeným přístupem nejsou otevřená data, hledí se na ně v kontextu VDF jako na otevřená data a musí tedy splňovat všechny požadované podmínky kladené na otevřená data včetně katalogizace v NKOD, s výjimkou upravených podmínek pro jejich užití. Pro úpravu platí, že v podmínkách užití dat musí být uvedena a zdůvodněna všechna stanovená omezení, musí být stanoveny podmínky pro získání souhlasu k užití a musí být uveden kontakt na publikující organizaci pro zažádání o přístup k publikovaným datovým sadám.

Udělování fyzického přístupu k takovým datovým sadám a správu a registraci povolených uživatelů plně spadá na publikující organizace.

Přístupy k datům VS

OVM (případně SPUÚ) při výkonu veřejné správy používají údaje vedené v systému základních registrů a údaje vedené o jednotlivých entitách v jiných agendových systémech, pokud mají dle zákona povinnost publikovat svoje údaje jiným agendám. Tyto údaje tvoří PPDF a jednotlivá OVM k nim mají přístup prostřednictvím referenčního rozhraní veřejné správy na základě svých práv a povinností zachycených v RPP.

OVM může pro svou práci vyžadovat i údaje nad rámec oprávnění definovaných v RPP, případně může potřebovat údaje, které PPDF neobsahuje. V takových případech může přistupovat k veřejně dostupným údajům prostřednictvím VDF a k získání potřebných dat využije vhodnou publikovanou sadu otevřených dat prostřednictvím NKOD.

Pokud taková sada neexistuje, požádá ohlašovatele odpovídající agendy o doplnění údajů v podobě otevřených dat do VDF.

Veřejnost má otevřený přístup ke všem otevřeným datům a tedy také k údajům ve VDF prostřednictvím NKOD a Portálu otevřených dat (POD).

Základní charakteristiky VDF

VDF obsahuje datové zdroje, které slouží jako doplňkový zdroj dat pro OVM v takových případech, kdy OVM nemají k použití takových dat přímo z PPDF přidělena práva v RPP, nebo taková data PPDF vůbec neobsahuje.

V konečném důsledku se tímto způsobem rozšiřuje sdílení údajů VS nad rámec zajišťovaný ISZR a eGSB v propojeném datovém fondu (viz. cíl 5.10 informační koncepce). Významným a hodnotným efektem VDF je také jeho zpřístupnění veřejnosti ve formátu otevřených dat v rámci možností a podmínek vymezených v zákonech u jednotlivých agend.

Data ve VDF musí být poskytovány v podobě otevřených dat a musí tak splňovat všechny požadavky kladené na otevřená data dle zákona č.106/1999 Sb. VDF je základním prostředkem a metodou pro sdílení veřejných informací mezi veřejnoprávními subjekty navzájem a pro sdílení veřejných údajů mezi veřejnoprávní a soukromoprávní sférou.

Pokud OVS využívá data z VDF, považuje je za správná a nemusí ověřovat jejich správnost.

Poskytovatel dat do VDF tudíž musí:

- garantovat správnost publikovaných údajů,
- garantovat kvalitu publikovaných údajů,
- garantovat aktuálnost publikovaných údajů,
- garantovat pravidelnost aktualizace publikovaných údajů,
- zajišťovat automatickou notifikaci všech změn zaregistrovaným zájemcům s využitím funkcionality POD.

Nástroje VDF

VDF vytvořený na základě principů otevřených dat vytváří virtuální distribuovaný datový prostor, ve kterém OVS sdílí s veřejností i mezi sebou navzájem veřejné údaje evidované v jimi spravovaných ISVS. Správa takového virtuálního prostředí a udržení kvality, konzistence a aktuálnosti publikovaných dat je založena na správě metadat publikovaných datových sad, popisu datových zdrojů a sémantickém mapování prvků struktur datových sad na sémantický slovník pojmů VS s pomocí k tomu určených nástrojů.

Správa dat VDF bude v základu probíhat stejným způsobem jako jsou již spravována otevřená data VS s několika dalšími rozšířeními. Základem bude katalog dat NKOD, který bude upraven a doplněn tak, aby mohl rozlišovat jednotlivé kategorie dat, umožňoval registrace uživatelů a zajišťoval notifikace změn.

Přehled všech nástrojů využitých při práci s daty VDF a otevřenými daty uvádí následující obrázek.

POD zastřešuje nástroje a funkční oblasti NKOD nutné pro katalogizaci a správu všech kategorií dat v PPDF.

Základní zajišťované funkce:

- katalogizace otevřených dat (VDF a OD),
- automatická notifikace změn v publikovaných datových sadách.

Základní logické celky POD:

- NKOD - upravený a doplněný pro správu informací o jednotlivých kategoriích dat a jejich zařazení do VDF.
  - NKOD je nástroj, ve kterém jednotlivé OVS katalogizují jimi zveřejňovaná otevřená data a jiné OVS a veřejnost v něm otevřená data vyhledává a získává k nim přístup.
  - Existence NKOD a povinnost jednotlivých OVS v něm katalogizovat svá otevřená data je dána zákonem č. 106/1999 Sb., o svobodném přístupu k informacím, který zavádí definici otevřených

dat a zakládá existenci NKOD.
- Katalog uživatelů otevřených dat (KUOD) - bude evidovat, jaké OVS využívají jaká otevřená data ve VDF.
- Katalog (registr) uživatelů dat s požadavkem notifikace změn v publikovaných datových sadách - seznam uživatelů s požadavky na zasílání notifikací o změnách pro konkrétní publikované datové sady.
- Notifikační HUB - nástroj zajišťující s pomocí KUOD s požadavkem notifikace automatickou distribuci notifikací změn v publikovaných sadách při ohlášení změny ze strany publikujícího OVS. Při registraci datové sady předává publikujícímu OVS URI pro zasílání informací o změnách.
- Registr požadovaných datových sad VDF - seznam požadavků na doplnění chybějících obrazů PPDF do VDF.

Sémantický slovník pojmů - slovník pojmů veřejné správy, jako nástroj harmonizace sémantiky otevřených dat, bude vytvořen na bázi výčtů údajů vedených k agendám v RPP dle § 51 odst. 5 písm. g), h) a i) zákona č. 111/2009 Sb. o základních registrech a bude je postupně rozpracovávat do podoby opakovaně použitelných a sdílených informačních modelů (ontologií), které budou propojeny na údaje vedené k agendám a též propojovány se slovníky a ontologiemi vznikajícími z iniciativy EU (např. ISA Core Vocabularies). Logická datová schémata popisující jejich strojové (syntaktické) vyjádření na logické úrovni budou propojena na pojmy sémantického slovníku pojmů, čímž bude realizováno propojení sémantiky (významu) dat napříč jednotlivými datovými sadami a jejich zdroji. Sémantický slovník pojmů bude nedílnou součástí VDF a stane se základem popisu sémantiky publikovaných dat a jejich vzájemného propojení.

Nástroj pro správu sémantického slovníku pojmů (NSSSP) – umožní správu sémantického slovníku pojmů všemi OVS a popisovat jejich data za účelem sémantické harmonizace.

Portál otevřených dat (POD) – vstupní brána do světa otevřených dat VS. Obsahuje NKOD, vzdělávací informace, publikační standardy, šablony, doporučené postupy, související dokumenty a informace o otevřených datech zveřejněných ve Standardech publikace a katalogizace otevřených dat

Lokální katalog otevřených dat - LKOD - volitelně implementován pro potřeby katalogizace otevřených dat konkrétního poskytovatele dat. NKOD pravidelně automaticky přebírá informace o zveřejněných datových sadách z jednotlivých lokálních katalogů.

Katalog aplikací nad otevřenými daty – seznam všech služeb (aplikací) využívajících publikované datové sady s možnostmi vyhledávání dle využívaných datových sad, nebo na základě životních situací konzumenta dat.

Obsahové rozdělení datových zdrojů VDF

VDF je tvořen dvěma druhy datových zdrojů:

- datové zdroje obsahující veřejný obraz PPDF,
- datové zdroje rozšiřující obraz PPDF o data, která v PPDF nejsou.

Veřejně přístupným obrazem datového zdroje chápeme data z tohoto zdroje transformovaná do strojově čitelné podoby, kterou lze zveřejnit. V případě, že lze datový zdroj zveřejnit celý v jeho původní podobě, je touto transformací identita. Tj. datový zdroj je v tomto případě sám sobě veřejně přístupným obrazem. U ostatních datových zdrojů transformace v maximální možné míře zachovává detaily původních dat a provádí tak typicky jen nezbytně nutnou anonymizaci. Ta většinou znamená buď projekci dat (tj. odstranění některých typů údajů z datové sady) nebo jejich statistickou agregaci.

Pro datové zdroje vytvářející veřejně přístupný obraz PPDF by měla platit následující jednoduchá obecná pravidla.

- Pro každý datový zdroj v PPDF musí existovat nějaký obraz ve VDF a to:

    - pro datové zdroje s legislativně stanoveným dálkovým přístupem budou předmětem publikace vždy primární data,
    - u datových zdrojů bez legislativně stanoveného dálkového přístupu budou v případě, že tomu nebude nic bránit, publikována primární data, v opačném případě bude jejich publikace provedena formou statistik a agregací s co nejjemnější granularitou.

Výčet datových zdrojů tvořící obsah VDF

- Obrazy základních registrů - jako obraz registru budou publikovány statistiky z registru
    - ROS
    - RÚIAN
    - RPP
    - ROB
- Obrazy veřejných rejstříků (veřejné rejstříky právnických a fyzických osob podle zákona č. 304/2013 Sb. (Zákon o veřejných rejstřících právnických a fyzických osob.).
    - spolkový rejstřík,
    - nadační rejstřík,
    - rejstřík ústavů,
    - rejstřík společenství vlastníků jednotek,
    - obchodní rejstřík,
    - rejstřík obecně prospěšných společností.
- Obrazy údajů evidovaných v RPP dle § 51 odst. 5 písm. h) zákona č. 111/2009 Sb. o základních registrech.
- Obrazy všech rejstříků, u kterých zákon definuje jejich dálkový přístup, pokud neexistují žádné právní překážky nebo významná rizika (v případě rejstříků bez stanoveného dálkového přístupu v zákonech - publikace agregací nebo statistik).
- Publikované (i lokální) číselníky, jejich sémantické provázání, případná konsolidace a uvedení kontextů zachycujících jejich použití.
- Publikované statistiky a jiné údaje využívané OVS, a které jsou vhodné pro veřejné zpřístupnění.

Zjednodušený koncept publikace dat do VDF

Cílem publikace dat VS do VDF je vytvoření harmonizovaného a konzistentního obrazu PPDF ve VDF, sémanticky provázaného i na odpovídající pojmy legislativy s pomocí Sémantického slovníku pojmů, včetně zachycených vzájemných souvislostí. Aby byla zajištěna konzistence VDF, a aby se z něj vytvořil rovnocenný obraz PPDF je proto nutné, aby publikované datové sady datových zdrojů v PPDF byly publikované i na úrovni propojených dat. Toho se dosáhne publikací konkrétních distribucí datových sad z PPDF se strojově čitelnou vazbou na sémantický slovník pojmů.

- Základem publikace do VDF je tedy sémantický slovník pojmů vytvořený na základě údajů v RPP, a který popisuje PPDF a tedy i budoucí obrazy ve VDF.
- Vlastní publikace bude realizovaná distribucemi datových sad s využitím takového formátu, který umožní zaznamenání souvislostí v datech a vazeb na sémantický slovník pojmů ve strojově čitelné podobě.

Publikované obrazy datových zdrojů PPDF ve VDF vytvoří infrastrukturu pro všechna otevřená data VS, což prakticky znamená, že publikující organizace nebudou muset při publikaci nových datových sad publikovat data, která již vypublikovaná budou, ale pouze budou publikovat nové rozšiřující informace k jednotlivým entitám VDF. Do jejich povinností ale přibude povinnost doplnit vazby nově publikovaných údajů na již existující entity ve VDF.

Publikovaná datová sada z datového zdroje PPDF musí k tomu, aby se stala součástí VDF:

- musí splňovat všechny podmínky kladené na otevřená data ve smyslu zákona č. 106/1999 Sb. o svobodném přístupu k informacím,
- musí být publikována dle předepsaných otevřených formálních norem ve smyslu § 4b odst. 1 zákona č.

106/1999 Sb. o svobodném přístupu k informacím
- musí mít doplněn kontext, který mapuje strukturální prvky datové sady na sémantický slovník pojmů.

Dále platí následující pravidla:

- V případě duplicit u publikovaných datových sad ve VDF, nebo doplňování údajů různými publikujícími ke stejné publikované entitě, spadá povinnost doplnění vazeb na již publikované údaje vždy na toho publikujícího, který publikuje doplňující údaje.
- U každé datové sady publikované do VDF musí být uvedena informace o notifikačním mechanismu o změnách dle příslušné OFN uvedené na POD.
- K zajištění právní závaznosti publikovaných dat musí být garantována jejich kvalita, aktuálnost a bezodkladná aktualizace, prostřednictvím organizačního a procesního zajištění v organizaci poskytovatele dat.

Správa dat VS ČR

Zapojení VDF do výkonu veřejné správy vyžaduje zajištěnou garanci kvality publikovaných datových sad, což znamená zajistit publikaci skutečně právně závazných, platných a pravidelně aktualizovaných datových sad s jasně definovanou zodpovědností OVS za takové sady. Je zřejmé, že takový stav bude obtížné dosáhnout stávajícími přístupy k publikaci otevřených dat ve VS, kdy publikace probíhá ne příliš koordinovaně, často nestandardními způsoby a s absencí používání doporučených standardů.

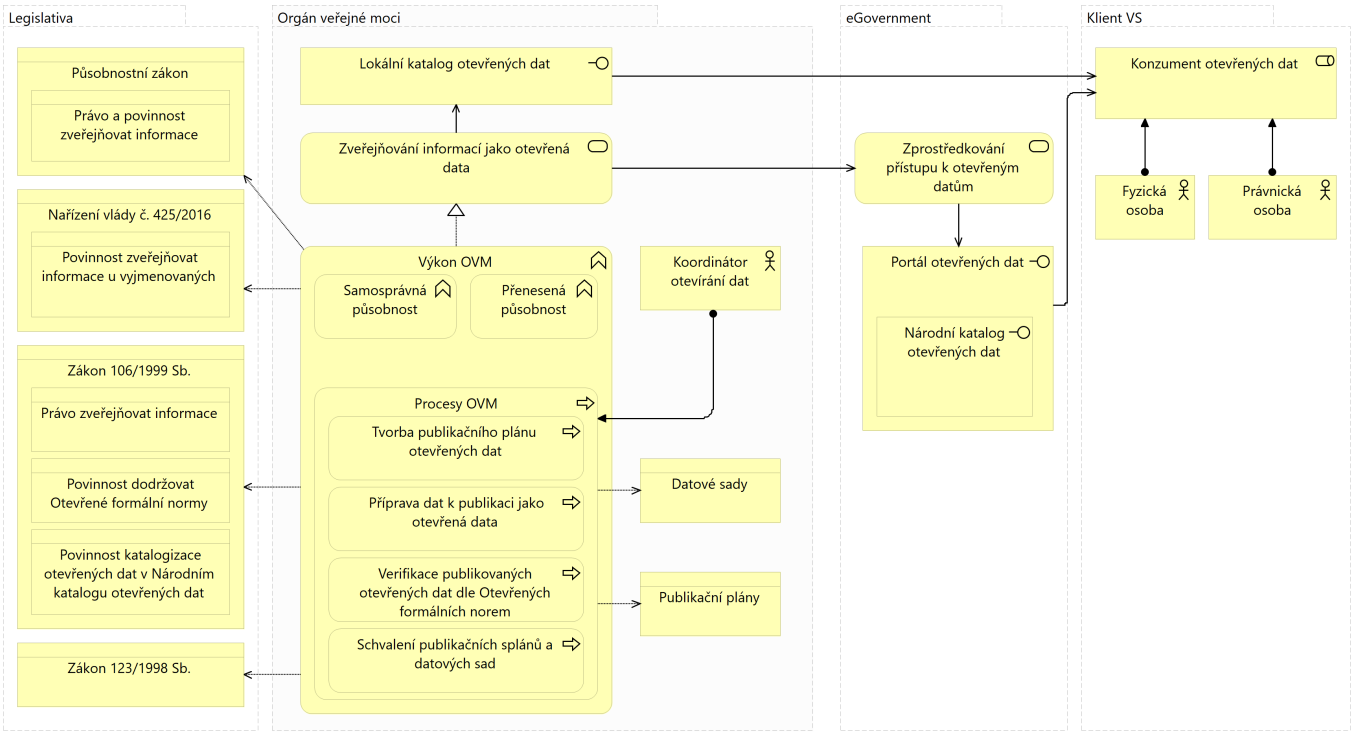
K zajištění kvalitního veřejného datového fondu bude nutné přistoupit k plánovanému a přesně definovanému vytváření VDF a také ke změně stávajícího paradigmatu správy dat VS ČR - oddělit data VS od aplikací a posunout je do centra řízení informatiky v souladu s postavením dat v legislativě. Správa a řízení dat VS se zdá být v budoucnu nezbytná i ve světle proklamací, že data jsou to nejcennější co organizace, a tedy i veřejná správa a v konečném důsledku i veřejnost, mají.

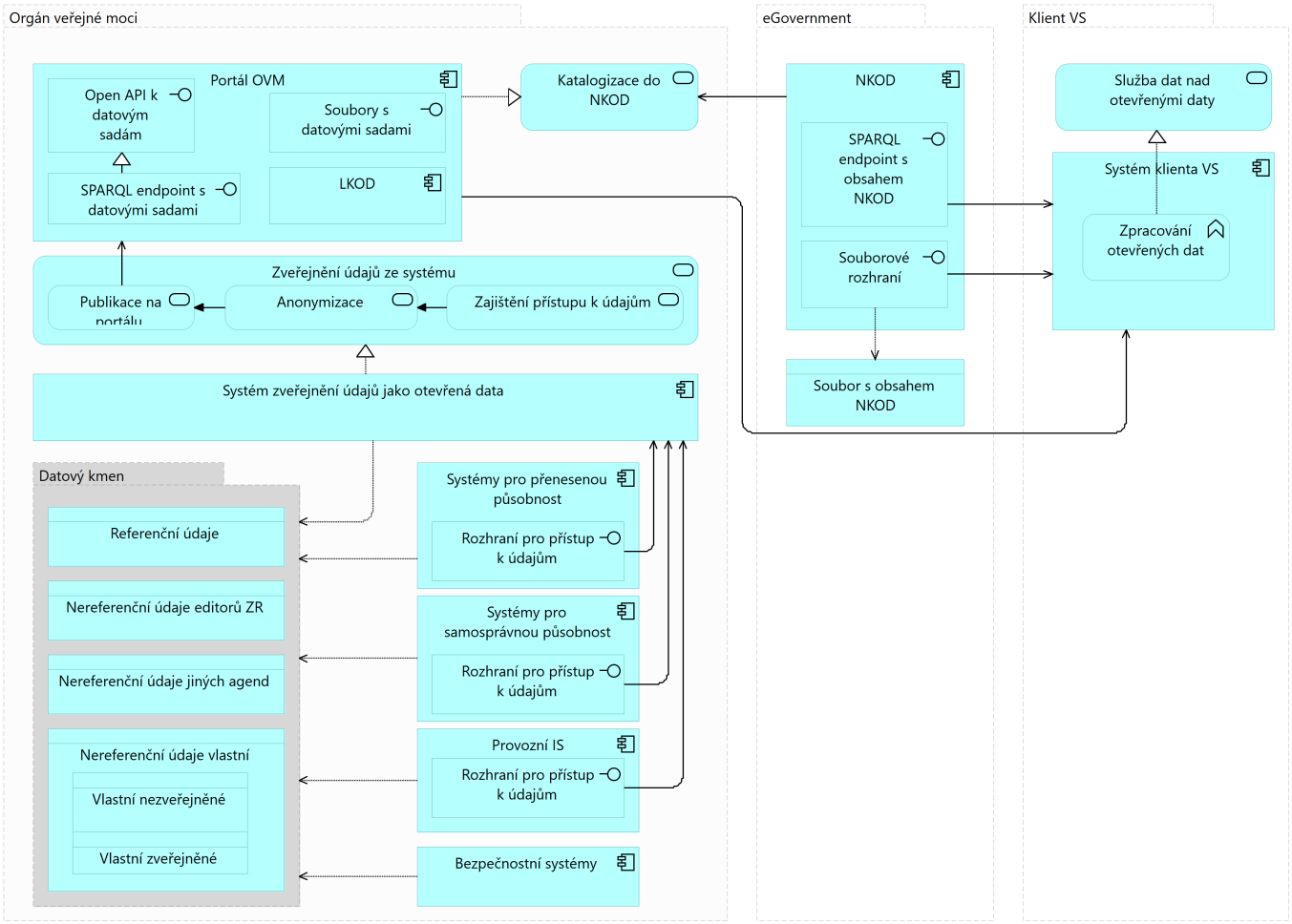
K zajištění kvality publikovaných datových sad, budou vydány:

- “Pravidla a postupy publikace dat do VDF”,
- “Pravidla a postupy publikace otevřených dat”,
- “Datová politika VS ČR”.

Dále budou k zajištění snadného využití publikovaných datových sad soustavně rozvíjeny stávající a vytvářeny nové OFN (Otevřené formální formy), určující koncepční i technické vzory pro vlastní publikaci dat. OFN jsou a budou publikovány na Portálu otevřených dat a jsou dle zákona č.106/1999 Sb. pro publikující organizace závazné.

Pohledy na veřejný datový fond





Otevřená data, Open Data, OD, VDF, Portál otevřených dat, NKOD, tematická oblast

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

Permanent link:
https://archi.gov.cz/nap:verejny_datovy_fond?rev=1570515595

Last update: **2019/10/08 08:19**