Digitální a informační agentura

Implementační průvodce datové sady (DATAHUB)

Tento dokument slouží jako **hlavní vstupní bod** pro pochopení datové sady. Pokud nevíte, kde začít, začněte zde.

1. Co je tato datová sada DATAHUB a proč existuje

Tato datová sada popisuje data integrovaná do datové platformy DATAHUB a slouží jako podklad pro analýzy a vizualizace v rámci projektu díky nimž můžeme poskytnout první celkový pohled na postup digitalizace veřejné správy v České republice.

Jedná se o NEagregovaná data obsahující jednotlivé transakce a případně jejich etapy z daného informačního systému.

Co je to transakce

Transakcí rozumíme jedno využití služby, kterou informační systém nabízí (rozumějte služby veřejné správy dle RPP). Jinými slovy by transakce dala popsat jako jedno podání/žádost ze strany občana, orgánu VS nebo firmy. Pokud transakce začíná podáním žádosti, pak je zakončena vyřízením žádosti a ukončením řízení. Víme, že služby veřejné správy jsou velmi různorodé, proto průběh, i charakter jednotlivých transakcí předpokládáme obdobně pestrý, proto jsme v sadě datových položek pro transakci zvolili maximálně obecné a srovnatelné údaje.

Co je to etapa

Etapa je dílčí částí transakce, je to její fáze. V některých případech se etapa může shodovat s úkonem definovaným v RPP, ovšem není to tak vždy, stejně tak může transakce procházet systémovými etapami, které vycházejí z architektury systému. Etapy nám pomáhají ve větším detailu nahlédnout na průběh transakce a umožní tak detailnější analýzu.

Proč tato data sbíráme

- Díky získávání transakčních dat můžeme **analyzovat využívání digitálních služeb**
- Řešíme tím problém absence **meziresortního pohledu na digitalizaci** v Česku
- Výsledky zpracování těchto dat slouží k informovanému rozhodování v rámci státní správy, vlády ČR a k prezentaci veřejnosti

Jak jsou data zabezpečena

Data, která nám posíláte, ukládáme v rámci uzavřené a zabezpečené infrastruktury DIA. Přístup k datovému úložišti bude přidělen pouze odpovědné osobě na vaší straně (poskytovatele dat) a to pouze pro zápis.

Poskytovatel dostane k přístupu ke své složce v Azuer storage unikátní a časově omezený SAS (Shared Access Signature) token, umožňující pouze upload souboru do složky. Zasílání souboru probíhá pomocí Azure storage REST-API rozhraní voláním PUT metody a za použití přístupového tokenu dojede k uploadu datového souboru do dedikovaného Azure storage containteru.

Formát souborů: XML, JSON nebo CSV.

V případě csv: separátor “,” (čárka), pole v dvojitých uvozovkách

Kódování: UTF-8

2. Kdy a v jaké situaci se data posílají

Data sbíráme **průběžně ideálně na denní bázi.** Součástí sběru dat je i zaslání iniciační datové dávky za období od 1.1.2024 do data zahájení datové výměny

Přijímáme jak ukončené tak neukončené transakce, v případě aktualizace dat umíme dle unikátního ID jednotlivé řádky aktualizovat.

Co tato datová sada není

- Nejde o statistické údaje, sbíráme surová transakční data.
- V žádném případě nechceme zpracovávat osobní údaje podléhající GDPR

3. Povinná a nepovinná data – co se stane, když chybí

Povinná data

- Základní údaje o každé transakci – ID, výsledek, identifikátory služby, IS, OVM, časové údaje začátku a konce, vstupní a výstupní kanály
- Jde o minimální datovou sadu.

Nepovinná data

- Zvyšují kvalitu výstupů
- Umožňují detailnější analýzu na míru danému ISVS, můžeme obohatit vaši analytiku pro vaše využití v rámci resortu.

4. Validace a technický kontrakt

Oficiální technický popis

1. JSON Schema: metriky_schema.json
2. XML Schema: metriky_schema.xsd
3. Excel katalog: Datový katalog DATAHUB V3.xlsx
4. Technické zadání: DATAHUB_TECHNICKE_ZADANI_PRO_POSKYTOVANI_DATOVYCH_SAD_2025_VER1_1.2

Jak si ověřit správnost

- validace proti schématu (JSON, XML) či datovému katalogu v případě CSV
- logická kontrola (např. časové vazby, unikátní ID apod.)

5. Nejčastější otázky (FAQ)

1. **Otázka:** Musíme mít přesně stejný dataset jako jiný úřad?

   Odpověď: Snažíme o maximální sjednocení dat, aby bylo možné je analyzovat společně a umožnit tak unikátní nadresortní pohled na českou digitalizaci.

2. **Otázka:** Co když některá data v systému nemáme?

   Odpověď: Chápeme, že každý systém je jiný a každá agenda má jiné náležitosti. Zároveň je nám zřejmé, že kapacity na straně úřadů jsou omezené. Snažili jsme se datovou sadu maximálně zjednodušit a zároveň jsme připraveni vyjít vstříc specifickým požadavkům jednotlivých poskytovatelů dat.

3. **Otázka:** Kdo nám práci na přípravě dat zaplatí?

   Odpověď: Bohužel není možné z naší strany pomoci s financováním prací, máme ale k dispozici kvalitní analytický tým, který s přípravou datové sady může pomoci.

4. **Otázka:** Jaké je zákonné zmocnění pro tento požadavek?

   Odpověď:

   Dle usnesení vlády ČR ze dne 27. dubna 2026 č. 259 vláda ukládá členům vlády a vedoucím ostatních ústředních orgánů státní správy poskytovat součinnost a zajistit průběžné systematické předávání datových sad z centrálních AIS do národního dashboardu digitalizace DATAHUB spravovaného Digitální a informační agenturou, a to v rozsahu nezbytném pro naplnění projektu vlády v oblasti digitalizace „Nástroj pro měření efektivity digitalizace“ (ve struktuře definované DIA).

   Dle § 5 odst. 2 písm. a) ZoISVS jsou orgány veřejné správy (dále jen "OVS") povinny s DIA spolupracovat při plnění jejích úkolů podle § 4 ZoISVS, tedy i předávat data. Technicky je projekt založen na předávání dat z agendových informačních systémů (AIS) jednotlivých orgánů veřejné moci do systému DATAHUB prostřednictvím standardizovaných referenčních rozhraní (§ 2 odst. 1 písm. i) ZoISVS). Přímé předávání dat do systému DATAHUB je právně opřeno o § 4 odst. 1 písm. a) ZoISVS.

   Dále je DATAHUB součástí schválené Informační koncepce vlády ČR článek 5 a 5.1. ze dne 22.10.2025

   - Vyhodnocování realizace IKČR se bude opírat o data a metriky agregované ze všech informačních systémů a digitálních služeb veřejné správy do jednoho centrálního místa (DATAHUB).
   - OVS každý nový nebo podstatně změněný proces či službu veřejné správy a každý nový nebo podstatně změněný informační systém na jejich podporu navrhuje tak, aby umožňoval měřit využívání, výkon, kvalitu a efektivitu všech agend a zpětnou vazbu uživatelů služeb VS, včetně sdílení a agregace těchto dat na jedno centrální místo (DATAHUB), aby bylo možné vytvářet přehledy a srovnání.

6. Jak postupovat, když si nejste jistí

1. Podívejte se na **ukázková data**

2. Zkontrolujte **JSON/XSD Schema**
3. Podívejte se do **Excel katalogu**
4. Pokud stále nejasné → kontaktujte DATAHUB

7. Detail položek (srozumitelný popis)

Každá položka odpovídá jedné hodnotě ve výstupních datech

Identifikátor/označení transakce

Povinnost	**ANO**
Co vyjadřuje	Jednoznačně identifikuje konkrétní transakci (využití služby).
Proč ji sbíráme	Unikátní identifikátor, primární klíč datové sady.
Kdy se vyplňuje	Vždy při vzniku transakce.
Jak má hodnota vypadat	string, unikátní v rámci systému.
Příklad	{ "id": "ABC-2024-001" }
Nejčastější chyby	znovupoužití ID, technický identifikátor bez stability.

Výsledek transakce

Povinnost	**ANO**
Co vyjadřuje	Zjednodušený výsledek transakce (úspěch / neúspěch).
Proč ji sbíráme	Základní vyhodnocení funkčnosti služby.
Kdy se vyplňuje	Po ukončení transakce.
Jak má hodnota vypadat	integer dle číselníku. (1 úspěšně/2 neúspěšně, NULL pro neukončené transakce)
Příklad	{ "vysledek": 1 }
Nejčastější chyby	Záměna s výsledným stavem.

Výsledný stav

Povinnost	**NE**
Co vyjadřuje	Konkrétní agendový stav po ukončení online části služby.
Proč ji sbíráme	Rozlišuje různé stavy transakcí v momentě zaslání datové sady. Upřesňuje výsledek transakce, je specifický pro jednotlivé systémy.
Kdy se vyplňuje	Pokud existuje smysluplný agendový stav.
Jak má hodnota vypadat	string, krátký kód.
Příklad	{ "stav": "zadost_prijata" }
Nejčastější chyby	technické nebo lidské popisy.

Uživatelská spokojenost

Povinnost	**NE**
Co vyjadřuje	Hodnocení spokojenosti uživatele se službou.
Proč ji sbíráme	Vyhodnocení kvality služeb.
Kdy se vyplňuje	Pokud je spokojenost zjišťována.
Jak má hodnota vypadat	integer (1–5 - známkování jako ve škole).
Příklad	{ "spokojenost": 2 }
Nejčastější chyby	dopočítávání bez dotazu na uživatele.

Vazba na orgán veřejné moci

Povinnost	**ANO**
Co vyjadřuje	Identifikaci OVM odpovědného za poskytnutí služby.
Proč ji sbíráme	Umožňuje agregaci podle OVM.
Kdy se vyplňuje	Vždy.
Jak má hodnota vypadat	Identifikátor OVM dle https://rpp-opendata.egon.gov.cz/odrpp/datovasada/ovm.json
Příklad	{ "ovm": 66002222 }
Nejčastější chyby	nejednotné kódy.

Vazba na službu veřejné správy

Povinnost	**ANO**
Co vyjadřuje	Konkrétní službu veřejné správy.
Proč ji sbíráme	Porovnání a vyhodnocení služeb.
Kdy se vyplňuje	Vždy při využití služby.
Jak má hodnota vypadat	Identifikátor služby dle RPP https://rpp-opendata.egon.gov.cz/odrpp/datovasada/sluzby.json
Příklad	{ "sluzba": "S41946" }
Nejčastější chyby	názvy místo ID.

Vazba na vstupní obslužný kanál

Povinnost	**ANO**
Co vyjadřuje	Kanál, kterým uživatel službu zahájil.
Proč ji sbíráme	Analýza využití kanálů.
Kdy se vyplňuje	Při zahájení transakce.
Jak má hodnota vypadat	Kód kanálu dle https://rpp-opendata.egon.gov.cz/odrpp/datovasada/cis_typ_obsluznych_kanalu.json
Příklad	{ "kanal_vstup": "portal" }
Nejčastější chyby	záměna vstupu a výstupu.

Vazba na výstupní obslužný kanál

Povinnost	**ANO**
Co vyjadřuje	Kanál, kterým byl proces ukončen.
Proč ji sbíráme	Porovnání vstupu a výstupu.
Kdy se vyplňuje	Vždy, v případě neukončené transakce se vyplňuje NULL, po ukončení transakce hodnota dle číselníku RPP, viz dále.
Jak má hodnota vypadat	Kód kanálu dle https://rpp-opendata.egon.gov.cz/odrpp/datovasada/cis_typ_obsluznych_kanalu.json
Příklad	{ "kanal_vystup": "datova_schranka" }
Nejčastější chyby	chybějící hodnota.

Čas zahájení

Povinnost	**ANO**
Co vyjadřuje	Okamžik zahájení transakce.
Proč ji sbíráme	Výpočet doby trvání, vytíženost systému
Kdy se vyplňuje	Vždy při zahájení.
Jak má hodnota vypadat	datetime (ISO 8601).
Příklad	{ "cas_zahajeni": "2024-01-01T10:00:00Z" }
Nejčastější chyby	Čas, který neodpovídá zahájení transakce

Čas ukončení

Povinnost	**ANO**
Co vyjadřuje	Okamžik ukončení transakce. V případě neukončené transakce zůstává NULL. Povinný jako element v datové struktuře, nepovinné naplnění hodnotou.
Proč ji sbíráme	Vyhodnocení délky transakce.
Kdy se vyplňuje	Po ukončení.
Jak má hodnota vypadat	datetime (ISO 8601).
Příklad	{ "cas_ukonceni": "2024-01-01T10:05:00Z" }
Nejčastější chyby	vyplnění před ukončením.

Sada volitelných parametrů

Povinnost	**NE**
Co vyjadřuje	Doplňující kontextové informace.
Proč ji sbíráme	Rozšíření analytických možností.
Kdy se vyplňuje	Pokud jsou data k dispozici.
Jak má hodnota vypadat	název s předponou vp_ libovolný datový typ
Příklad	{"vp_": { "vp_vydany_dokument": "Výpis z Registru obyvatel"}}
Nejčastější chyby	ukládání povinných dat.

Informační systém

Povinnost	**ANO**
Co vyjadřuje	Systém, ve kterém transakce proběhla.
Proč ji sbíráme	Porovnání IS.
Kdy se vyplňuje	Vždy.
Jak má hodnota vypadat	Identifikátor IS dle https://rpp-opendata.egon.gov.cz/odrpp/datovasada/isvs.json
Příklad	{ "is": "89" }
Nejčastější chyby	Hodnota mimo číselník

Zdrojový informační systém

Povinnost	**NE**
Co vyjadřuje	IS, odkud data pochází.
Proč ji sbíráme	Sledování toků dat.
Kdy se vyplňuje	Pokud existuje zdroj–cíl.
Jak má hodnota vypadat	Identifikátor IS dle https://rpp-opendata.egon.gov.cz/odrpp/datovasada/isvs.json
Příklad	{ "is_zdroj": "89" }
Nejčastější chyby	shoda s cílovým IS bez důvodu.

Cílový informační systém

Povinnost	**NE**
Co vyjadřuje	IS, do kterého data směřují.
Proč ji sbíráme	Analýza integrací.
Kdy se vyplňuje	Pokud existuje cílový IS.
Jak má hodnota vypadat	Identifikátor IS dle https://rpp-opendata.egon.gov.cz/odrpp/datovasada/isvs.json
Příklad	{ "is_cil": "89" }
Nejčastější chyby	vyplnění bez integrace.

Velikost požadavku

Povinnost	**NE**
Co vyjadřuje	Objem přenášených dat.
Proč ji sbíráme	Kapacitní a výkonové analýzy.
Kdy se vyplňuje	Pokud je relevantní.
Jak má hodnota vypadat	integer (např. v bajtech).
Příklad	{ "bytes": 2048 }
Nejčastější chyby	odhad místo skutečné hodnoty.

Místní příslušnost

Povinnost	**NE**
Co vyjadřuje	Zda je služba využívána v místní příslušnosti a jakého typu
Proč ji sbíráme	Analýzy využívání místní příslušnosti dle typů
Kdy se vyplňuje	Pokud je služba realizována v rámci místní příslušnosti
Jak má hodnota vypadat	typ místní příslušnosti
Příklad	{ "mistni_prislusnost": "ze zákona" }
Nejčastější chyby	uvedení názvu města/úřadu místo typu místní příslušnosti

Vykonávající úřad

Povinnost	**NE**
Co vyjadřuje	Konkrétní úřad vykonávající službu.
Proč ji sbíráme	Detailní reporting, regionální analýzy
Kdy se vyplňuje	Pokud je znám a je odlišný od OVM poskytujícího data
Jak má hodnota vypadat	Identifikátor úřadu (IČO, ID datové schránky nebo ID OVM)
Příklad	{ "urad": "123445" }
Nejčastější chyby	duplicita s OVM.

Identifikátor/označení etapy

Povinnost	**ANO (pokud jsou etapy uváděny)**
Co vyjadřuje	Jednoznačně identifikuje etapu transakce.
Proč ji sbíráme	Analýza průběhu transakce.
Kdy se vyplňuje	Při členění transakce na etapy.
Jak má hodnota vypadat	string.
Příklad	{ "id": "E1" }
Nejčastější chyby	chybějící vazba na transakci.

Název etapy

Povinnost	**ANO (pokud jsou etapy uváděny)**
Co vyjadřuje	Lidsky čitelný název etapy.
Proč ji sbíráme	Srozumitelnost výstupů.
Kdy se vyplňuje	Pokud existuje etapa.
Jak má hodnota vypadat	string.
Příklad	{ "nazev": "Podání žádosti" }
Nejčastější chyby	technické názvy.

Vazba etapy na úkon služby veřejné správy

Povinnost	**NE**
Co vyjadřuje	Úkon dle Katalogu služeb
Proč ji sbíráme	Provázání s RPP.
Kdy se vyplňuje	Pokud existuje odpovídající úkon.
Jak má hodnota vypadat	Identifikátor úkonu.
Příklad	{ "ukon": "U123" }
Nejčastější chyby	volný text.

Etapa Kanál

Povinnost	**ANO (pokud jsou etapy uváděny)**
Co vyjadřuje	Obecné označení použitého kanálu.
Proč ji sbíráme	Přehled využití kanálů.
Kdy se vyplňuje	Při každé transakci.
Jak má hodnota vypadat	Kód kanálu.
Příklad	{ "kanal": "portal" }
Nejčastější chyby	nekonzistentní hodnoty.

Etapa čas

Povinnost	**ANO (pokud jsou etapy uváděny)**
Co vyjadřuje	Časový údaj, kdy došlo k zahájení etapy.
Proč ji sbíráme	Proměření délky etap a průběhu celé transakce
Kdy se vyplňuje	U každé etapy
Jak má hodnota vypadat	Datum a čas
Příklad	{ "cas": "2024-01-01T10:05:00Z " }
Nejčastější chyby	nekonzistentní časové údaje

Etapa sada volitelných parametrů

Povinnost	**NE**
Co vyjadřuje	Doplňující kontextové informace k etapě.
Proč ji sbíráme	Rozšíření analytických možností.
Kdy se vyplňuje	Pokud jsou data k dispozici.
Jak má hodnota vypadat	název s předponou vp_ libovolný datový typ
Příklad	{"vp_": { "vp_vydany_dokument": "Výpis z Registru obyvatel"}}
Nejčastější chyby	ukládání povinných dat.

Financováno z projektu: Vytvoření centrální datové infrastruktury pro práci s daty rč CZ.31.5.0/0.0/0.0/24_130/0010992