# Ako používať stránku na sťahovanie údajov o kvalite ovzdušia

**Dátum: Marec 2025**

## História dokumentu

| **Verzia** | **Dátum** | **Autor(i)** | **Poznámky** |
| --- | --- | --- | --- |
| 0.1 | 2023-06-29 | AE |  |
| 0.2 | 2023-09-08 | AE |  |
| 0.3 | 2023-11-14 | AE |  |
| 0.4 | 2024-04-15 | AE |  |
| 0.5 | 2024-05-03 | AE |  |
| 0.6 | 2024-06-06 | AE |  |
| 0.7 | 2024-11-18 | AE |  |
| 0.8 | 2025-03-31 | MV |  |

Obsah

[Ako používať stránku na sťahovanie údajov o kvalite ovzdušia 1](#_Toc199319985)

[História dokumentu 1](#_Toc199319986)

[Prehľad 3](#_Toc199319987)

[1. Stránka služby na sťahovanie údajov o kvalite ovzdušia 4](#_Toc199319988)

[1.1. Filtre 4](#_Toc199319989)

[1.2. Stiahnutie údajov 7](#_Toc199319990)

[1.3. Metaúdaje 7](#_Toc199319991)

[1.4. Slovník pojmov 8](#_Toc199319992)

[1.5. Dokumentácia 9](#_Toc199319993)

[2. API pre službu sťahovania údajov o kvalite ovzdušia 10](#_Toc199319994)

[2.1. Príklady skriptov 13](#_Toc199319995)

[2.2. Vizualizácia Parquet súborov 17](#_Toc199319996)

[3. Definícia schémy Parquet 17](#_Toc199319997)

## ****Prehľad****

Tieto pokyny vysvetľujú, ako používať novú službu sťahovania údajov o kvalite ovzdušia prostredníctvom webovej aplikácie aj API.

* **Webová aplikácia:** <https://eeadmz1-downloads-webapp.azurewebsites.net>
* **API:** <https://eeadmz1-downloads-api-appservice.azurewebsites.net/swagger/index.html>

V dokumente sú podrobne predstavené rôzne možnosti dostupné v rámci webovej aplikácie na stiahnutie údajov, ako aj príklady použitia API na sťahovanie bez potreby použitia webovej aplikácie.

Údaje je možné stiahnuť v dvoch rôznych formátoch:

1. **Súbory Parquet:** Obsahujú kompletné časové rady príslušných odberových miest.
2. **URL adresy:** Vygeneruje sa CSV súbor s URL adresami, ktoré odkazujú na súbory Parquet pre každé požadované odberové miesto.

## ****1. Stránka služby na sťahovanie údajov o kvalite ovzdušia****

Stránka na sťahovanie je rozdelená na dva hlavné bloky:

1. **Rozhranie:** Táto funkcia umožňuje filtrovať a vybrať požadované údaje na stiahnutie priamo do vášho prehliadača (na ľavej strane).
2. **Popis služby sťahovania:** Podrobné informácie o tejto službe sú dostupné na pravej strane obrazovky.



Okrem toho stránka poskytuje informácie o metaúdajoch, odkaz na zoznam slovníkov, ktoré popisujú definície v stiahnutých súboroch, a odkaz na túto dokumentáciu.

### ****1.1. Filtre****

V tejto časti sa vysvetľuje, ako fungujú filtre v aplikácii. Dôležité je vedieť, že **žiadny z filtrov nie je povinný** – ak nie je filter vyplnený, sťahovanie sa vykoná pre všetky možné hodnoty.

**Príklad:** Ak je filter „znečisťujúce látky“ prázdny, budú stiahnuté súbory so všetkými znečisťujúcimi látkami dostupnými v systéme.

#### Krajiny

Pomocou tohto filtra je možné vybrať jednu alebo viac konkrétnych krajín, pre ktoré sa majú údaje stiahnuť. Ak nie je vybraná žiadna krajina, budú stiahnuté údaje pre všetky krajiny.

#### Mestá

Ak vybrané krajiny obsahujú mestá, zobrazí sa nový filter s príslušnými mestami. Rovnako ako pri filtroch krajín je možné vybrať jedno alebo viac miest, alebo nechať pole prázdne na získanie údajov pre všetky mestá.

Geometrické informácie o každom meste pochádzajú z údajov Urban Audit. Pomocou týchto informácií a súradníc staníc (dostupných v odkaze) bola vykonaná priestorová analýza na priradenie staníc jednotlivým mestám.

#### Znečisťujúce látky

Je možné filtrovať informácie podľa požadovaných znečisťujúcich látok. Môžete vybrať jednu alebo viac možností, alebo nechať pole prázdne a tým zahrnúť všetky znečisťujúce látky.

#### Dataset

Tu je možné vybrať:

* **Overené údaje (E1a)** – nahlásené krajinami.
* **Najnovšie neoverené údaje (Up-To-Date / UTD / E2a)** – priebežne prenášané.
* **Historické údaje Airbase (2002 – 2012).**

Tento filter **je povinný**.

#### Typ

Určuje, či sa údaje zbierajú v **hodinových**, **denných** alebo **variabilných** intervaloch (napr. týždenné, mesačné a pod.). Možno vybrať konkrétnu možnosť, alebo ponechať všetko.

**Poznámka:** Tento filter sa neuplatňuje pri žiadostiach typu „Zoznam URL“ alebo „Zhrnutie“.

#### E-mail

Voliteľné pole. Odporúča sa zadať, aby sa v prípade problémov dala ľahšie zabezpečiť spätná väzba používateľovi.

#### ****Časové pokrytie (Temporal coverage)****

Je možné filtrovať a stiahnuť údaje za konkrétne časové obdobie. Na tento účel je potrebné zvoliť **začiatok (Start)** a **koniec (End)** obdobia, ktoré sa má stiahnuť.

V prípade, že bol vybraný dataset **Up-To-Date (E2a)**, je možné zvoliť typické časové pokrytie pre stiahnutie.

Ak nie je zadané žiadne časové pokrytie, budú stiahnuté **všetky dostupné údaje**.

**Poznámka:** Tento filter **nemá vplyv na požiadavky typu „Zoznam URL“ alebo „Zhrnutie“**.
Zhrnutie berie do úvahy celé časové rady podľa miesta a datasetu, zatiaľ čo Zoznam URL poskytuje odkazy na celé súbory bez ohľadu na časové krytie a typ agregácie.

#### ****URL adresy****

V závislosti od potreby je možné stiahnuť **iba URL adresy** k vybraným súborom namiesto samotných súborov vo formáte Parquet.

Na tento účel stačí **zaškrtnúť príslušné políčko**.

Ak je táto možnosť aktivovaná, výsledok **nebude brať do úvahy filtre týkajúce sa času alebo typu agregácie** – bude obsahovať len odkazy na celé súbory.

### ****1.2. Stiahnutie údajov****

Po výbere požadovaných filtrov je potrebné kliknúť na tlačidlo **„Download“ (Stiahnuť)**. Po krátkom čakaní sa súbory automaticky stiahnu cez prehliadač.

Namiesto stiahnutia súborov je tiež možné kliknúť na tlačidlo **„Summary“ (Zhrnutie)**, ktoré poskytne prehľad o počte súborov a veľkosti sťahovaných údajov. Tento náhľad je užitočný na odhadovanie objemu údajov. Ak sa prekročí **maximálny limit 300 MB**, odporúča sa namiesto údajov stiahnuť **URL adresy**, ktoré vedú na jednotlivé súbory.

Ako už bolo spomenuté vyššie, na získanie URL adries treba aktivovať možnosť **„List of URLs“** (Zoznam URL).



### ****1.3. Metaúdaje****

Kliknutím na odkaz **„Metadata“** sa otvorí nová stránka so zhrnutím všetkých metaúdajov zo všetkých krajín. Nájdete tu interaktívnu tabuľku s informáciami o meraniach kvality ovzdušia hlásených v rámci **AQ e-Reporting** do repozitára **Eionet CDR**.

🔗 **Priamy odkaz:**
<https://discomap.eea.europa.eu/App/AQViewer/index.html?fqn=Airquality_Dissem.b2g.measurements>

Na tejto stránke je možné sťahovať údaje a filtrovať podľa viacerých možností. Treba si však uvedomiť, že tu je možné získať **iba metaúdaje** (o sieťach, staniciach, odberových bodoch atď.), **nie samotné údaje o znečisťujúcich látkach** – na to slúži stránka sťahovania popísaná v tomto dokumente.



### ****1.4. Slovník pojmov****

Odkaz **„Vocabulary“ (Slovník)** vedie na stránku so zhrnutím rôznych slovníkov pojmov, ktoré sa používajú v súboroch Parquet. Kliknutím na odkazy sa používateľ dostane do **Eionet Data Dictionary**, kde sú definované jednotlivé pojmy.( [https://dd.eionet.europa.eu/)](https://dd.eionet.europa.eu/%29)

Aj tu platí, že stránka poskytuje iba **metaúdaje**, **nie merania znečisťujúcich látok**. Na sťahovanie meraní je potrebné použiť hlavnú stránku služby sťahovania.



### ****1.5. Dokumentácia****

Cez odkaz **„Documentation“ (Dokumentácia)** sa otvorí nový panel v prehliadači s týmto dokumentom.

## 2. API pre službu sťahovania údajov o kvalite ovzdušia

Existuje možnosť prístupu k API, kde je možné nájsť dostupné metódy súvisiace so sťahovaním súborov.

🔗 **API odkaz:**
<https://eeadmz1-downloads-api-appservice.azurewebsites.net/swagger/index.html>



Na tejto stránke si môžete vybrať požadovanú metódu.
Na vykonanie požiadavky je potrebné kliknúť na tlačidlo **„Try it out“** pri príslušnom endpointe.

**Najdôležitejšie endpointy:**

* **ParquetFile:** Stiahne ZIP súbor so všetkými filtrovanými Parquet súbormi.
* **ParquetFile/async:** Stiahnutie na pozadí. Endpoint vráti URL adresu, na ktorej sa súbor vygeneruje.
* **ParquetFile/urls:** Stiahne CSV súbor s URL adresami zodpovedajúcimi filtrovaným Parquet súborom.

Všetky tieto endpointy používajú rovnaké parametre na filtrovanie a získanie požadovaných údajov.



Endpoint **ParquetFile/urls** **ignoruje** filtre pre dátum a typ agregácie.

Po vyplnení požadovaných hodnôt stačí kliknúť na **„Execute“**.

Parametre je potrebné vyplniť nasledovne:

* **countries:** Kódy krajín oddelené čiarkou. Ak nie je zadané nič (prázdny zoznam), filter sa neuplatní a stiahnu sa údaje pre všetky krajiny.
*Príklad:* ["IS", "DK", "NO"]
* **cities:** Názvy miest oddelené čiarkou. Ak nie je zadané nič, stiahnu sa údaje pre všetky mestá.
*Príklad:* ["Berlin", "Madrid"]
* **pollutants:** Zadajte buď úplné URL adresy k jednotlivým znečisťujúcim látkam (získané kliknutím na odkazy v slovníku), alebo ich označenie.
*Príklad:*
["http://dd.eionet.europa.eu/vocabulary/aq/pollutant/3", "http://dd.eionet.europa.eu/vocabulary/aq/pollutant/1"]
alebo
["SO2", "CO"]



* **dataset:** Hodnota datasetu:
	1. Neoverené údaje priebežne prenášané (Up-To-Date / UTD / E2a) – od 2023
	2. Overené údaje (E1a) – z rokov 2013–2022, nahlásené do 30. septembra za predchádzajúci rok
	3. Historické údaje Airbase – z rokov 2002–2012
* **source (voliteľné):** Označuje, odkiaľ bola požiadavka vykonaná:
	1. "website" – ak bola použitá webová stránka
	2. "API" alebo "Custom script" – ak bola použitá API priamo
* **dateTimeStart:** Dátum a čas, od ktorého sa majú údaje filtrovať.
*Formát:* rrrr-mm-ddTHH:MM:SSZ
*Príklad:* 2024-05-27T12:00:19Z
*(Tento parameter sa ignoruje pri ParquetFile/urls)*
* **dateTimeEnd:** Dátum a čas do ktorého sa majú údaje filtrovať.
*Formát:* rrrr-mm-ddTHH:MM:SSZ
*Príklad:* 2024-05-28T12:00:19Z
*(Tento parameter sa ignoruje pri ParquetFile/urls)*
* **aggregationType:** Určuje typ údajov:
	1. Hodinové (hour)
	2. Denné (day)
	3. Variabilné intervaly (napr. týždenné, mesačné)

*(Ignorované pri ParquetFile/urls)*

* **email (voliteľné):** Na identifikáciu používateľa, čo umožní jednoduchšiu komunikáciu v prípade problémov.

Na konci kliknite na **„Download file“**, čím získate ZIP súbor s Parquet údajmi.



Endpoint **DownloadSummary** ignoruje dátumové a agregačné filtre, aj keď sú zadané.

## 2.1. Príklady skriptov

**Príklad Python skriptu – základný**

Nasleduje príklad Python skriptu na stiahnutie Parquet súborov z Azure so zvolenými filtrami.

**Poznámka:** Cesta downloadPath a názov súboru fileName sa nastavujú podľa toho, kam si používateľ želá súbor uložiť.

python

CopyEdit

import requests

apiUrl = "https://eeadmz1-downloads-api-appservice.azurewebsites.net/"

endpoint = "ParquetFile"

downloadPath = "localPath\\"

fileName = "download\_data.zip"

# Telo požiadavky

request\_body = {

 "countries": ["ES"],

 "cities": ["Madrid"],

 "pollutants": ["SO2"],

 "dataset": 1,

 "dateTimeStart": "2024-11-17T09:37:00.000Z",

 "dateTimeEnd": "2024-11-18T09:37:00.000Z",

 "aggregationType": "hour",

 "email": "exampleEmail@gmail.com"

}

# Odoslanie požiadavky

downloadFile = requests.post(apiUrl + endpoint, json=request\_body).content

# Uloženie súboru

with open(downloadPath + fileName, 'wb') as output:

 output.write(downloadFile)

**Príklad Python skriptu – async endpoint**

Tento príklad používa asynchrónny endpoint, ktorý je vhodnejší pre väčšie požiadavky.
Prvá požiadavka spustí proces a vráti URL adresu, kde bude výsledok dostupný. Skript následne čaká, kým sa výstup sprístupní.

python

CopyEdit

from datetime import datetime

import time

import requests

apiUrl = "https://eeadmz1-downloads-api-appservice.azurewebsites.net/"

endpoint = "ParquetFile/async"

downloadPath = ""

fileName = "EEA\_2023.zip"

request\_body = {

 "countries": ["IT"],

 "cities": [],

 "pollutants": ["NO2", "NOX as NO2", "PM10"],

 "dataset": 2,

 "dateTimeStart": "2023-01-01T00:00:00Z",

 "dateTimeEnd": "2023-01-05T23:59:59Z",

 "aggregationType": "day",

 "email": "my.own@email.com"

}

response = requests.post(f"{apiUrl}{endpoint}", json=request\_body)

downloadFile = response.text

print(downloadFile)

t\_start = datetime.now()

while True:

 if (datetime.now() - t\_start).total\_seconds() > 3600: # zastaví sa po 1 hodine

 break

 parquetResponse = requests.get(downloadFile)

 if parquetResponse.status\_code == 404:

 time.sleep(20)

 else:

 break

with open(downloadPath + fileName, "wb") as fp:

 fp.write(parquetResponse.content)

**Príklad Python skriptu – zoznam URL adries**

Tretí variant skriptu využíva endpoint na získanie **zoznamu URL adries**. Tento zoznam je vhodný na **veľké požiadavky alebo vlastné spracovanie dát**, pretože si používateľ môže údaje stiahnuť a spracovať vo svojom prostredí. Ide o najstabilnejší spôsob sťahovania.

python

CopyEdit

import os

import requests

apiUrl = "https://eeadmz1-downloads-api-appservice.azurewebsites.net/"

endpoint = "ParquetFile/urls"

downloadFolder = "./e1a"

request\_body = {

 "countries": ["IT"],

 "cities": [],

 "pollutants": ["NO2", "NOX as NO2", "PM10"],

 "dataset": 2,

 "dateTimeStart": "2023-01-01T00:00:00Z",

 "dateTimeEnd": "2023-01-05T23:59:59Z",

 "aggregationType": "hour",

 "email": "my.own@email.com"

}

response = requests.post(f"{apiUrl}{endpoint}", json=request\_body)

urls = response.text.split("\n")[1:]

os.makedirs(downloadFolder, exist\_ok=True)

for url in urls:

 fileName = url.split("/")[-1]

 with open(f"{downloadFolder}/{fileName}", "wb") as fp:

 fp.write(requests.get(url).content)

**Príklad použitia v Postmane**

V tomto príklade sa popisuje, ako získať Parquet súbory cez aplikáciu **Postman**.

1. Otvorte Postman.
2. Nastavte požiadavku ako **POST** a vložte telo požiadavky vo formáte **JSON** s požadovanými údajmi.
3. Po vykonaní požiadavky kliknite na **„Save response to file“**, čím uložíte výsledný ZIP súbor s Parquet údajmi.

{

„countries“: [„SK“],

„cities“:[‚Bratislava‘]

„pollutants“:[‚SO2‘]

...

}

## 2.2. Vizualizácia Parquet súborov

Existuje viacero online nástrojov, ktoré umožňujú **zobraziť obsah Parquet súborov** alebo ich **konvertovať do formátu CSV** tak, aby bolo možné ich čítať bez špecializovaných nástrojov.

Jeden z takýchto nástrojov je napríklad **Parquet Viewer**.

# 3. Definícia schémy Parquet

Parquet súbory obsahujú najdôležitejšie informácie nahlásené jednotlivými krajinami. Každý súbor obsahuje nasledovné atribúty:

* **Samplingpoint:** Identifikátor známy ako „lokálne ID“ odberového bodu. Má prefix krajiny. Tento identifikátor je jedinečný pre každú stanicu a vytvárajú ho národní reportéri, preto neexistuje univerzálny slovník pre tento parameter.
* **Pollutant:** Identifikátor znečisťujúcej látky.
*Viac informácií o znečisťujúcich látkach nájdete v príslušnom slovníku.*
* **Start:** Začiatok časového intervalu, v ktorom boli údaje zaznamenané.
*Formát:* rrrr-mm-dd H:M:S
* **End:** Koniec časového intervalu, v ktorom boli údaje zaznamenané.
*Formát:* rrrr-mm-dd H:M:S
* **Value:** Číselná hodnota merania pre daný znečisťujúci prvok v danom intervale.
* **Unit:** Jednotka, v ktorej je hodnota merania vyjadrená.
*Viac informácií o jednotkách nájdete v slovníku.*
* **AggType:** Typ agregácie – určuje, či sa údaje zbierali **hodinovo, denne alebo s premenlivým intervalom** (napr. týždenne, mesačne).
*Viac informácií nájdete v časti o typoch agregácie.*
* **Validity:** Udáva platnosť merania v rámci daného časového intervalu.
*Viac informácií nájdete v časti o platnosti údajov.*
* **Verification:** Stav overenia daného merania pre znečisťujúcu látku v danom čase.
*Viac informácií nájdete v časti o overení údajov.*
* **ResultTime:** Dátum a čas, kedy bol vytvorený súbor s reportovanými údajmi.
* **DataCapture:** Percento času, za ktoré bol odber vzoriek alebo pozorovanie vykonané.
* **FkObservationLog:** Tento stĺpec **nemá význam pre koncového používateľa** – ide o technickú informáciu pre interné účely.