



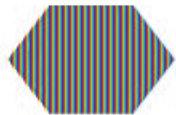
Dataverse API használata az ARP-ben

2025.06.25.

Pallinger Péter - HUN-REN SZTAKI

Mi az az API?

- Application Programming Interface
- Alkalmazásprogramozási felület
- <https://en.wikipedia.org/wiki/API>
- Számítógépes programok közötti kapcsolat megteremtésére szolgáló felület
- Minden program API-kon keresztül “beszélget” más programokkal
 - Pl. Openstack-nek (HUN-REN Cloud) is van API-ja



Dataverse API

- Hálózati, RESTful API
- Dataverse API dokumentáció:
<https://guides.dataverse.org/en/6.1/api/index.html>
- Minden, a hagyományos felhasználói felületen elérhető funkció elérhető API-n keresztül is.



Mi az a REST API

- Representational State Transfer
- <https://en.wikipedia.org/wiki/REST>
- HTTP alapú:
 - GET, POST, PUT, DELETE, stb. metódusokat használ
- Valamilyen strukturált adatformátummal
 - XML
 - **JSON**



JSON adatformátum

- JavaScript Object Notation

Példa:

```
{  
  "first_name": "John",  
  "last_name": "Smith",  
  "age": 27,  
  "address": {  
    "street_address": "21 2nd Street",  
    "city": "New York",  
    "state": "NY",  
    "postal_code": "10021-3100"  
  },  
  "phone_numbers": [  
    {  
      "type": "home",  
      "number": "212 555-1234"  
    },  
    {  
      "type": "office",  
      "number": "646 555-4567"  
    }  
  ]  
}
```

```
"phone_numbers": [  
  {  
    "type": "home",  
    "number": "212 555-1234"  
  },  
  {  
    "type": "office",  
    "number": "646 555-4567"  
  }  
]
```



REST API alapok 1

- HTTP kérések végeznek műveleteket, a HTTP metódusnak megfelelően. Például:
 - olvasás: GET

```
$ curl -s -X GET https://repo.researchdata.hu/api/datasets/3 | json_pp
```

```
{  
  "data" : {  
    "authority" : "21.15109",  
    "id" : 3,  
    "identifier" : "CONCORDA/XRAOAS",  
    "latestVersion" : {  
      .....  
    }  
  }  
}
```



REST API alapok 2

- létrehozás: POST

```
$ echo '{"datasetVersion": { "metadataBlocks": { "citation": {  
    "fields": [ { "value": [ { "authorName": { "value": "Próba, Baba",  
        "typeClass": "primitive", "multiple": false, "typeName":  
"authorName" } } ],  
        "typeClass": "compound", "multiple": true, "typeName": "author" }  
    ], "displayName": "Citation Metadata" } } } }' \  
| curl -s -X POST --upload-file - -H "X-Dataverse-key:$API_TOKEN"  
https://repo.researchdata.hu/api/dataverses/sztaki/datasets?doNotValidate=true
```



REST API alapok 3

- módosítás: PUT / PATCH

- `curl -H "X-Dataverse-key: $API_TOKEN" -X PUT https://repo.researchdata.hu/api/datasets/3 --upload-file dataset-update-metadata.json`

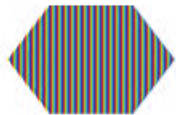
- törlés: DELETE

- `curl -X DELETE https://repo.researchdata.hu/api/datasets/3 -H "X-Dataverse-key: $API_TOKEN"`



Dataverse API-k

- Search API
- Data access API
- Native API
- SWORD API



Search API

- Adatcsomagok keresésére szolgál

```
$ curl -s "https://repo.researchdata.hu/api/search?q=tree&sort=date" | json_pp
{
  "data" : {
    "count_in_response" : 3,
    "items" : [
      {
        "authors" : [
          "Schmera, Dénes",
          "Baur, Bruno"
        ],
        "citation" : "Schmera, Dénes; Baur, Bruno, 2024, \"Catch data of the box tree moth in the suburb of Basel and data on the condition of trees in the nearby natural box-tree forest\", https://hdl.handle.net/21.15109/ARP/LACXB9, ARP, V1",
        "citationHtml" : "Schmera, D&eacute;nes; Baur, Bruno, 2024, \"Catch data of the box tree moth in the suburb of Basel and data on the condition of trees in the nearby natural box-tree forest\", <a href=\"https://hdl.handle.net/21.15109/ARP/LACXB9\" target=\"_blank\">https://hdl.handle.net/21.15109/ARP/LACXB9</a>, ARP, V1",
```



Data Access API

- Fájlok letöltése egy adatcsomagból
 - ZIP formátumban, korlátozott méretig

```
$ curl -L -O -J  
https://repo.researchdata.hu/api/access/dataset/:persistentId/?persistentId=hdl:21.15109/CORDA/4IQJ3W
```

```
$ ls -l dataverse_files.zip
```

```
-rw-r--r--  1 user user 1292762 jún   18 15.21 dataverse_files.zip
```

- Fájlrészek letöltésére

```
$ curl -H "range:bytes=89-157" $SERVER_URL/api/access/datafile/11
```

```
113.0    "Agabus affinis Ad."      12.0
```

```
114.0    "Agabus bipustulatus Ad."    12.0
```



Data Access API - további funkciók

- Fájlmetaadatok lekérése
- Több fájl letöltése
 - `/api/access/datafiles/$id1,$id2,...$idN`
 - Szintén zip-ben, méretkorláttal
- Hozzáférés kérelmezése
- Hozzáférési kérelmek
listázása/megadása/megtagadása
- Hozzáférés visszavonása



Native API

- Adatcsomagok (dataset), tárolók (dataverse collection) és fájlok kezelése
 - lekérdezés, létrehozás, törlés, módosítás, közzététel, visszavonás, provenance
- Jogosultságok kezelése
- Sémák (metadata block-ok) kezelése
- Kvóták kezelése
- Embargók kezelése
- Adminisztrációs funkciók



Klienskönyvtárak

- Számptalan nyelven léteznek
 - C/C++
 - Go
 - Java
 - Javascript
 - Julia
 - Python
 - R
 - Ruby



Esettanulmány

- 12000+ tároló létrehozása
- 70000+ adatcsomag feltöltése
- 40+TB
- párhuzamosítva
- ellenőrzéssel
- <https://gist.github.com/pallinger/dfc19457ab4ad5e2fff38f6d46950835>



Hasznos eszközök

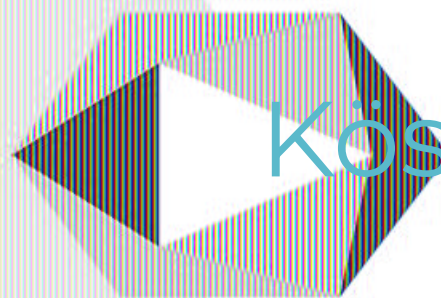
- DataverseMassUploader (python)
 - <https://github.com/dsd-sztaki-hu/DataverseMassUploader/>
- DVUploader (java)
 - <https://github.com/GlobalDataverseCommunityConsortium/dataverse-uploader>
- python-dvuploader (python)
 - <https://github.com/gdcc/python-dvuploader>
- Teszt JSON-ok
 - https://github.com/gdcc/dataverse_testdata



Gyakorlat

- Mindenki vegye elő a saját pitonját! (Vagy akár mást is (PHP, Ruby, R)!)
- Vagy:
 - arpoktatas.dsd.sztaki.hu
 - arpguest / *****





Köszönöm a figyelmet!

Kapcsolat:

Pallinger Péter

HUN-REN SZTAKI

pallinger.peter@sztaki.hu

