



Több száz gigabájtos kutatási adatcsomag repozitálása a HUN-REN ARP-ben

Ki, mit, hova, miért?

Trendek és innovációk a kutatási adat-kezelésben

Barancsuk Lilla

2025.10.08.

Bemutakozás: HUN-REN Energiatudományi Kutatóközpont a KFKI telephelyen

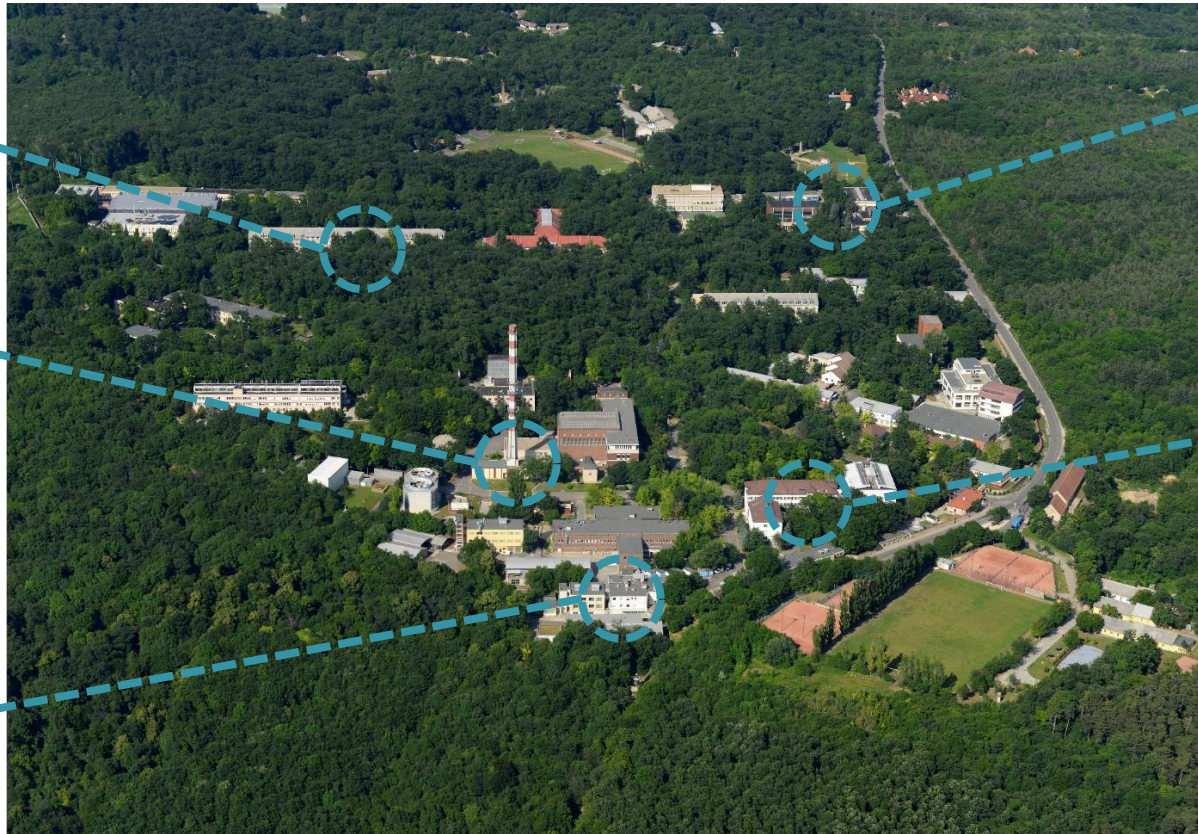
Anyagtudomány

Nukleáris energetika

Környezetvédelem

Űrkutatás

Energiatárolók és
hidrogéngazdaság



A Központi Fizikai Kutatóintézet csillebérci telephelye

A kutatócsoport

A meteorológia és az energetika határterületén



Napelemes termeléselőrejelzés



Felhőátvonulások hálózati hatásai



Kisfeszültségű elosztóhálózat
megújuló befogadóképessége



Energiaközösségek
modellezése

Környezetfizikai laboratórium

*Energia Stratégia és Környezeti
Hatások Kutatócsoport*

**Energetikai meteorológia és klimatológia
fókuszterület**



Sinkovics Bálint
Villamosmérnök
Tudományos munkatárs



Oláhné Groma
Veronika
Meteorológus,
Kutatócsoport-vezető



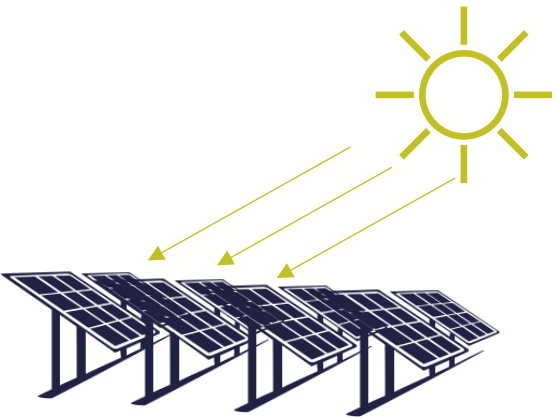
Günter Dalma
*Mérnök- informatikus,
doktorandusz*



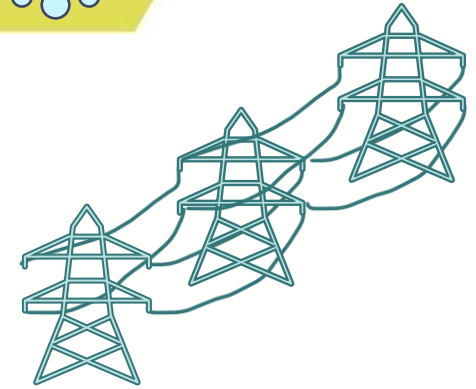
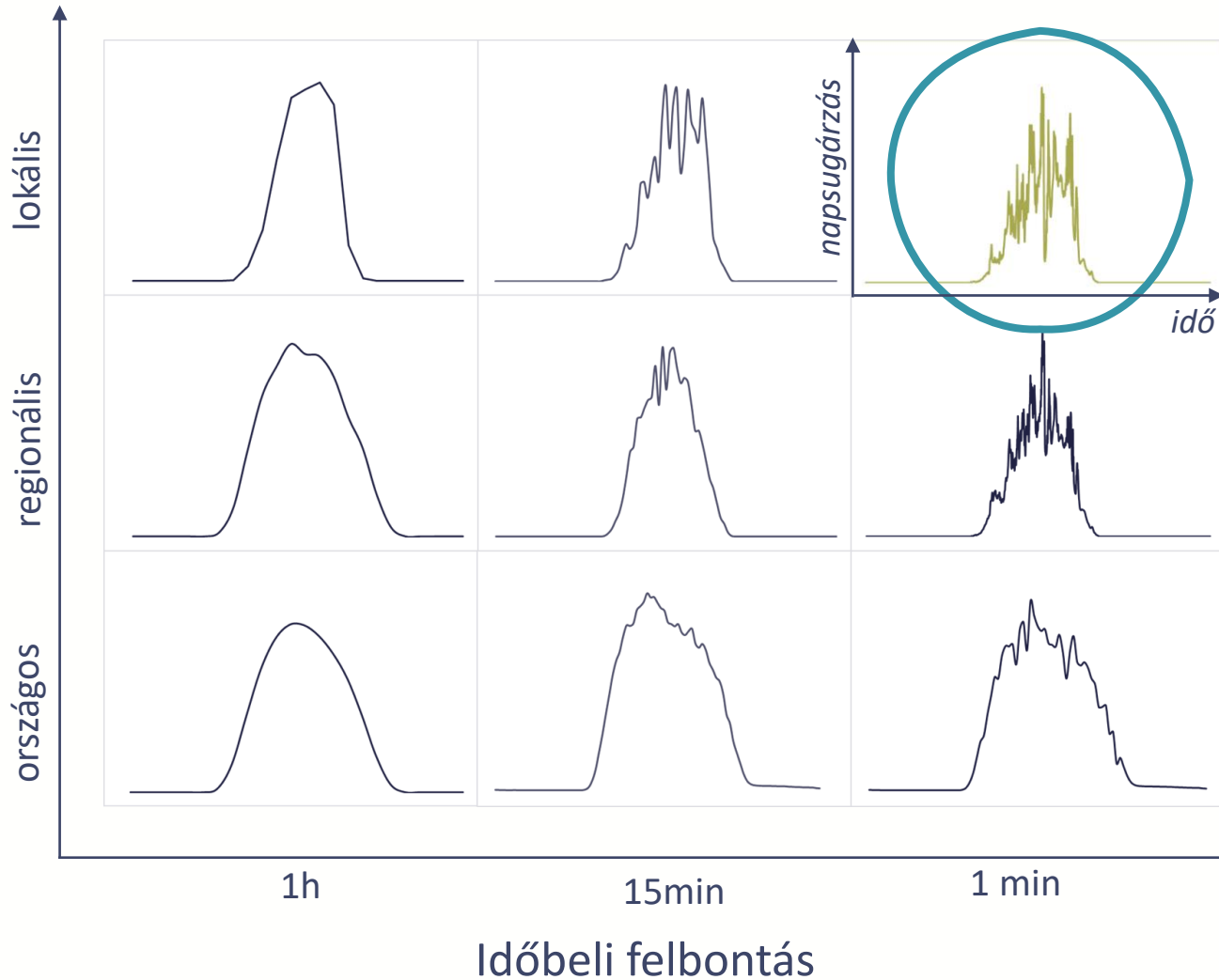
Barancsik Lilla
Villamosmérnök,
Doktorandusz
ARP Nagykövet

Ultrarövidtávú napelemtermelés-előrejelzés

Miért jó ez?



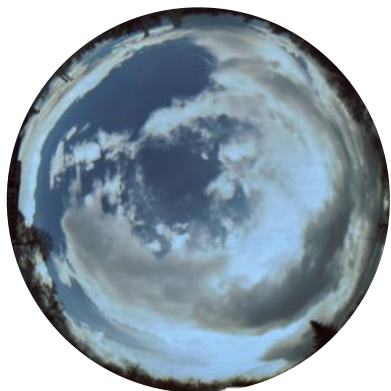
Térbeli felbontás



Égboltkamerás előrejelzés

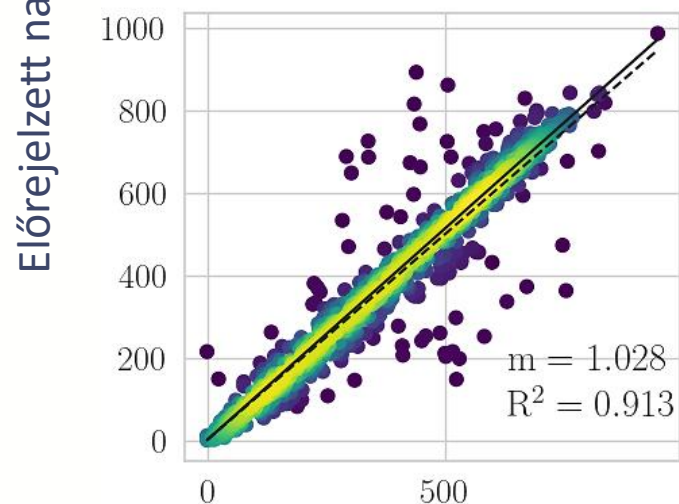
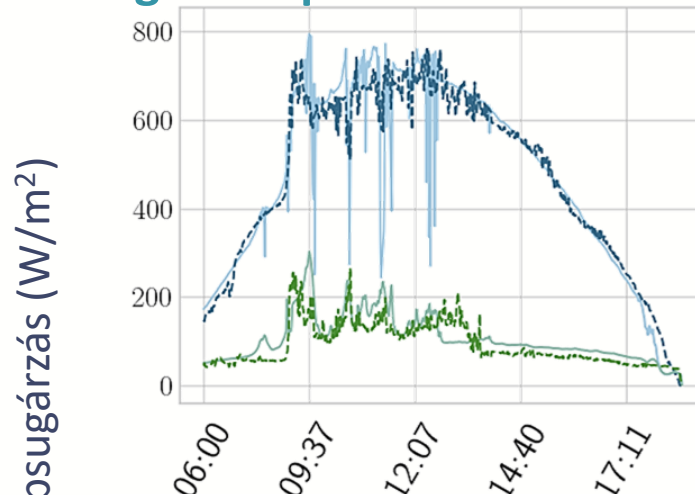
Képek → Napelemtermelés

Felhős nyári nap:

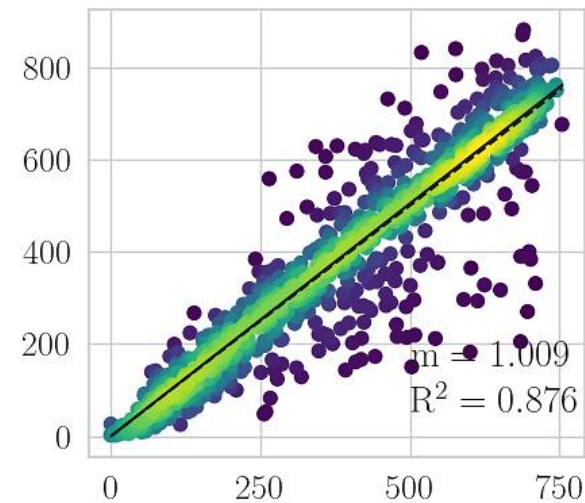
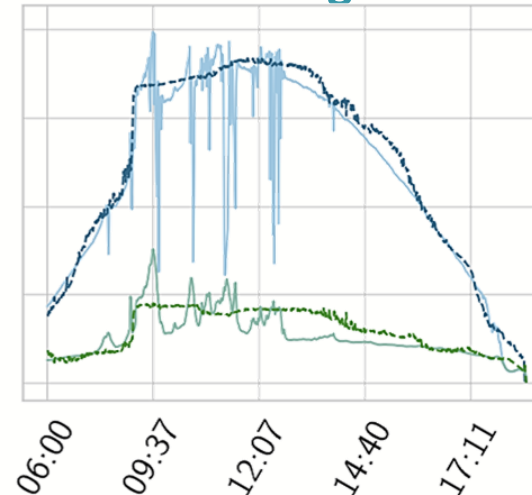


■ Globálsugárzás
■ Diffúz sugárzás

Égboltképek felhasználásával

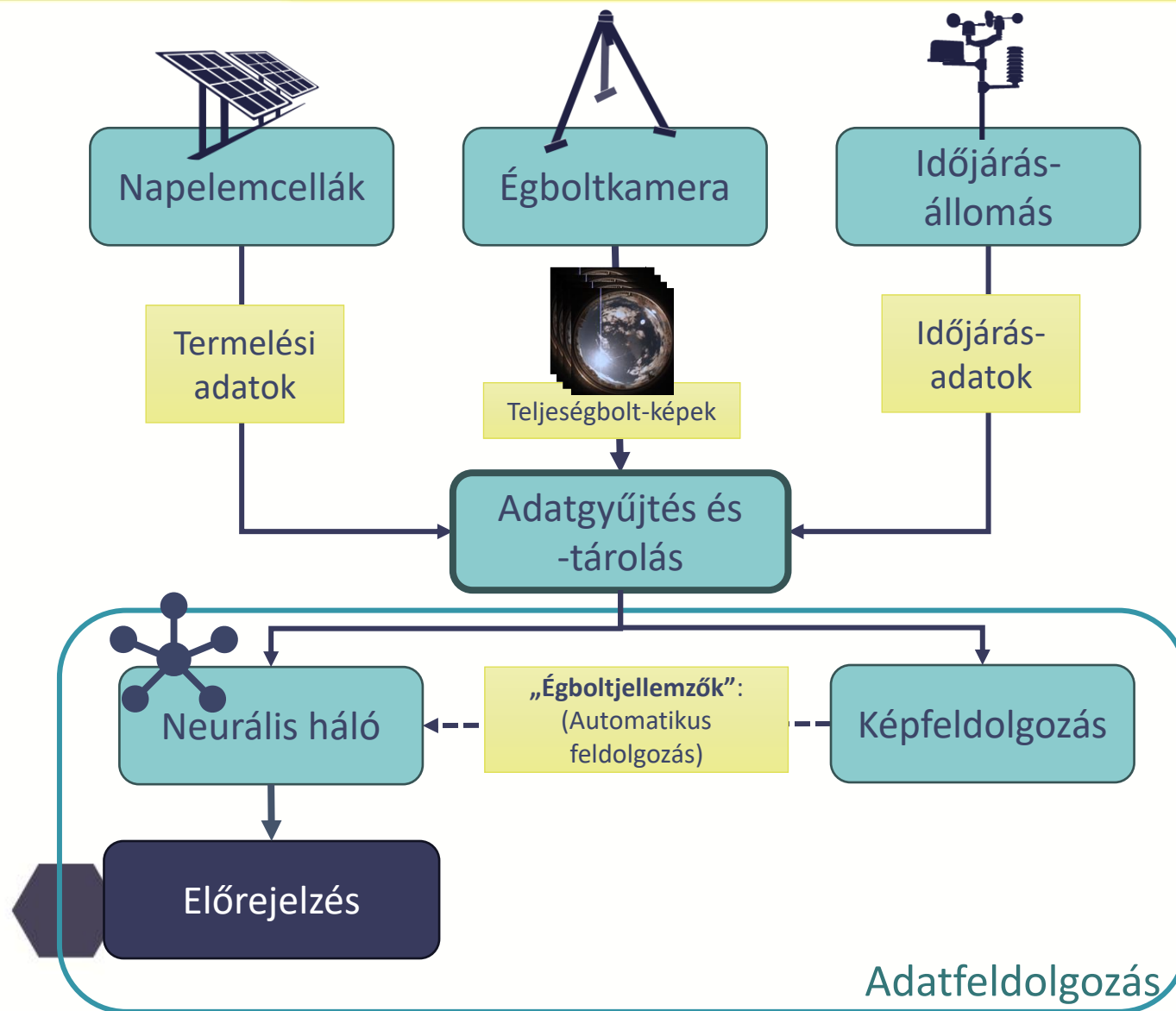


Csak meteorológiai mérések

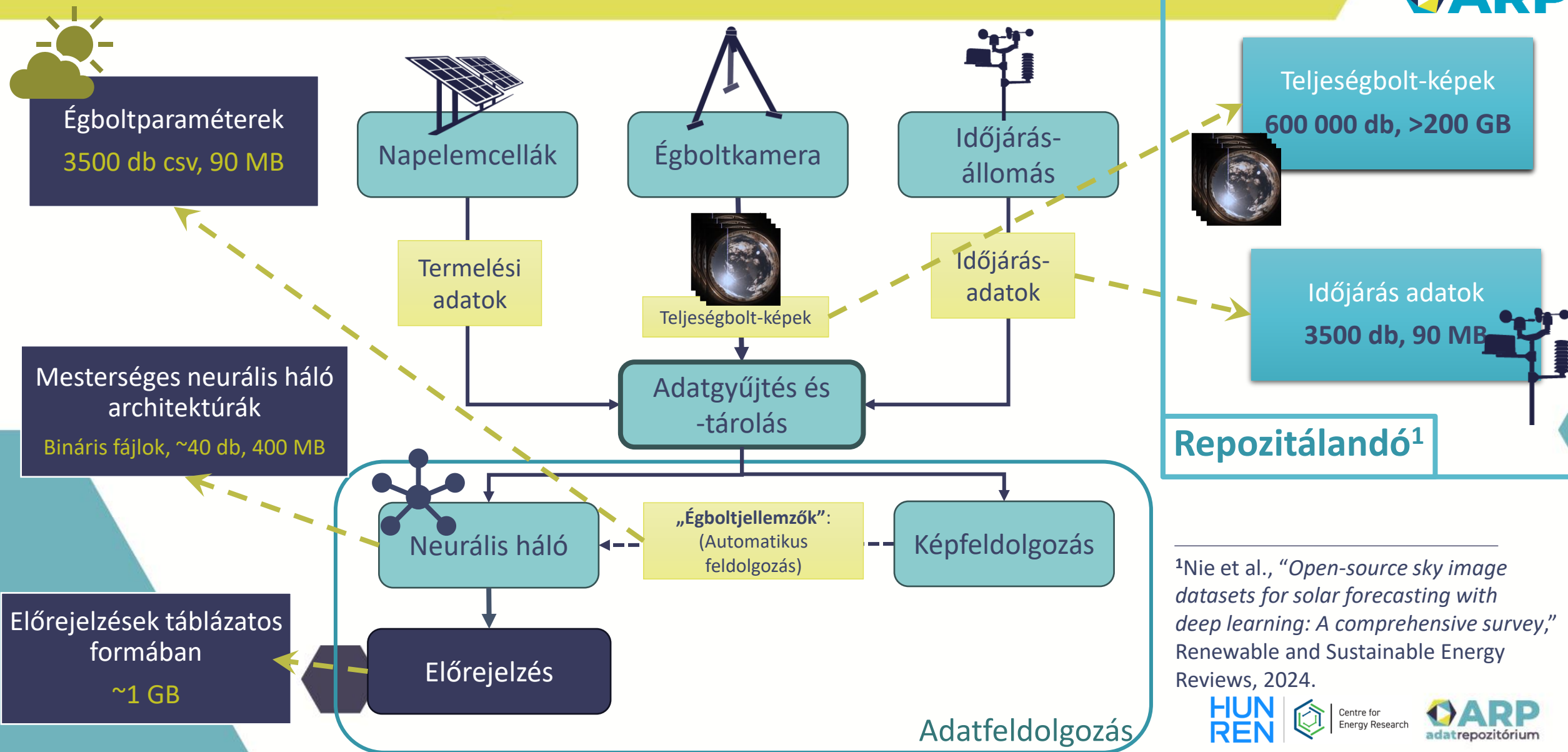


Mért napsugárzás (W/m²)

Milyen adatok szükségesek egy előrejelzéshez?



Milyen adatok szükségesek az előrejelzéshez?



¹Nie et al., “Open-source sky image datasets for solar forecasting with deep learning: A comprehensive survey,” Renewable and Sustainable Energy Reviews, 2024.

Repozitálási próbálkozások...



- 1 Kézi feltöltés (~10 GB .zip fájlok) → megakad feltöltés közben
- 2 Python script a képek egyéni feltöltéséhez → megakad feltöltés közben
- 3 Python script .zip-ek feltöltése → ARP percekig tölti az adatkészletet
- 4 To be continued...



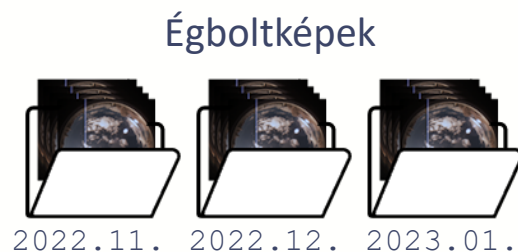
LOADING



LOADING



LOADING



Python Kliens

IOIO
IOIO



ARP
Adatrepozitórium

The
Dataverse[®]
Project

A Dataverse nem tud több tízezer fájlt
egy adatkészletben kezelni!

HUN
REN



Centre for
Energy Research

ARP
adatrepozitórium

A megoldás...

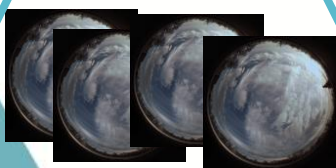
HUN-REN Energiatudományi Kutatóközpont Dataverse

Energiastratégia és Környezeti Hatások Kutatócsoport Dataverse

Új Dataverse: Ground-based all-sky imagery...



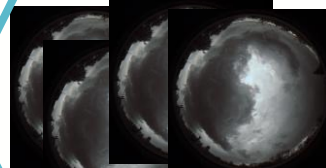
Adatkészletek



2022.11.

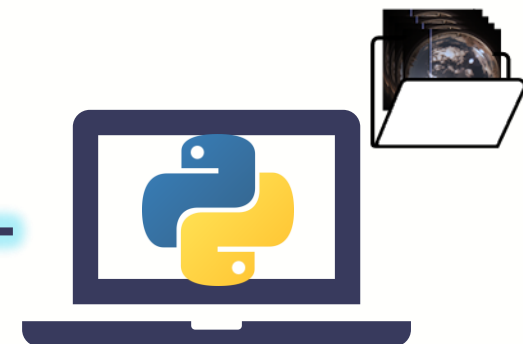


2022.12.



2023.01.

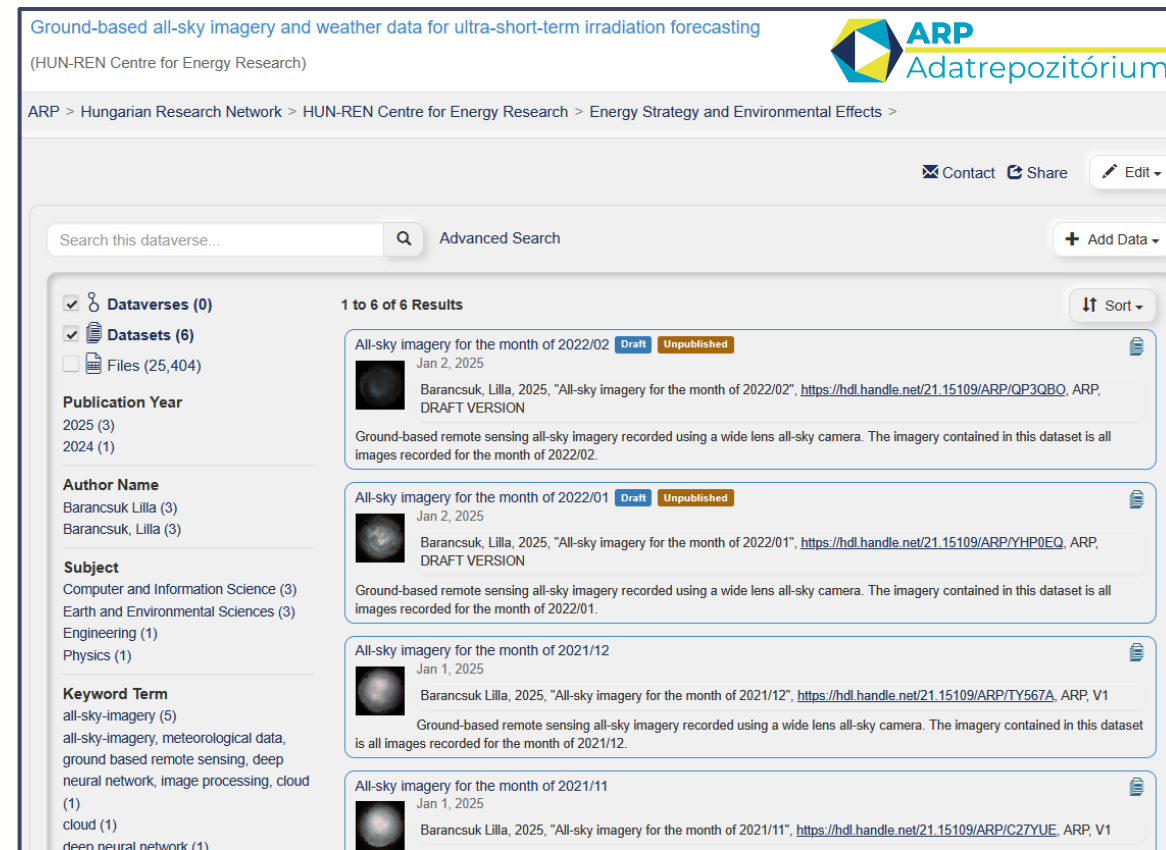
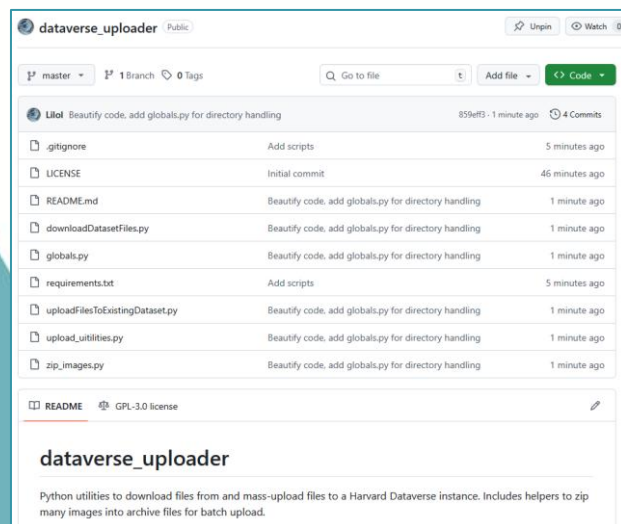
- Hónapok → egy-egy .zip fájlban
- **Fájlszintű metaadatoláshoz:** hónapok szerint külön adatkészlet a képekhez
- A teljes adatsor egyben is letölthető



Az adatkészlet: All-sky-imagery

- ✓ FAIR elvek
- ✓ Összesen 2 hónapnyi szabadonálló kép feltöltve
- ✓ *A többinél megakadt a feltöltés...*
- ✓ Dataverse kliens library-k érdeklődőknek:
<https://guides.dataverse.org/en/6.1/api/client-libraries.html>
- ✓ Személyes kedvencem Pythonosoknak: **pyDataverse**
<https://github.com/gdcc/pyDataverse>
- ✓ A feltöltő script forráskódja a GitHubon:

https://github.com/Lilol/dataverse_uploader



<https://repo.researchdata.hu/dataverse/all-sky-imagery>

Tanulságok



Dataverse jelenlegi architektúrája nem alkalmas több tízezer fájl egy adatkészletben kezelésére



Ilyenkor a többszintű struktúra jó megoldás lehet



Fontos az együttműködés fejlesztők, kutatók és adatgazdászok között



Az ARP csapatától nagyon sok **támogatást** kaptam

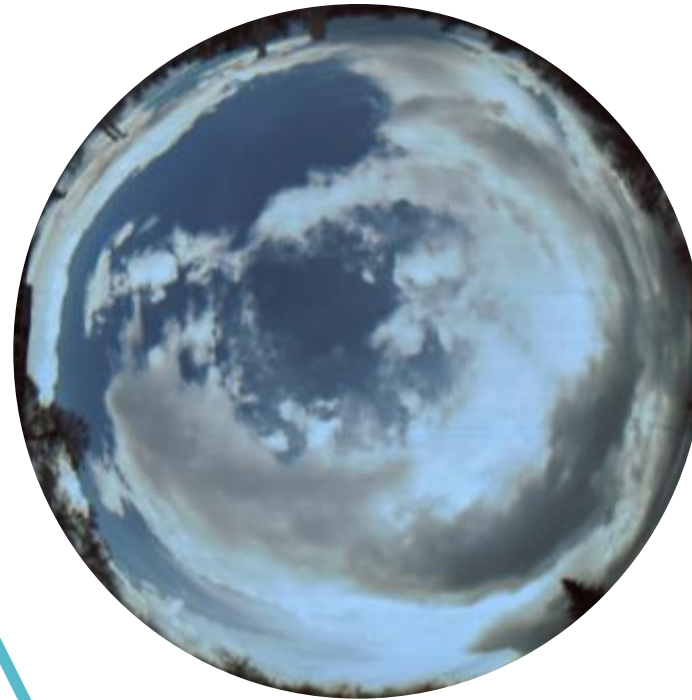


Tovább lépés: fájl szintű metaadatolás

Ha Te is nagy adatcsomaggal küzdesz...
Szívesen segítek a repozitálásban:
✉ barancsuk.lilla@ek.hun-ren.hu



<https://repo.researchdata.hu/dataverse/all-sky-imagery>



LOADING

Köszönöm a
figyelmet!



https://github.com/Lilol/dataverse_uploader

