

# Adatgyűjtési rendszerek specifikációja és tervezése

---

Mező György

HUN-REN Csillagászati és Földtudományi Kutatóközpont

NUN-REN Wigner Fizikai Kutatóközpont Adatközpont

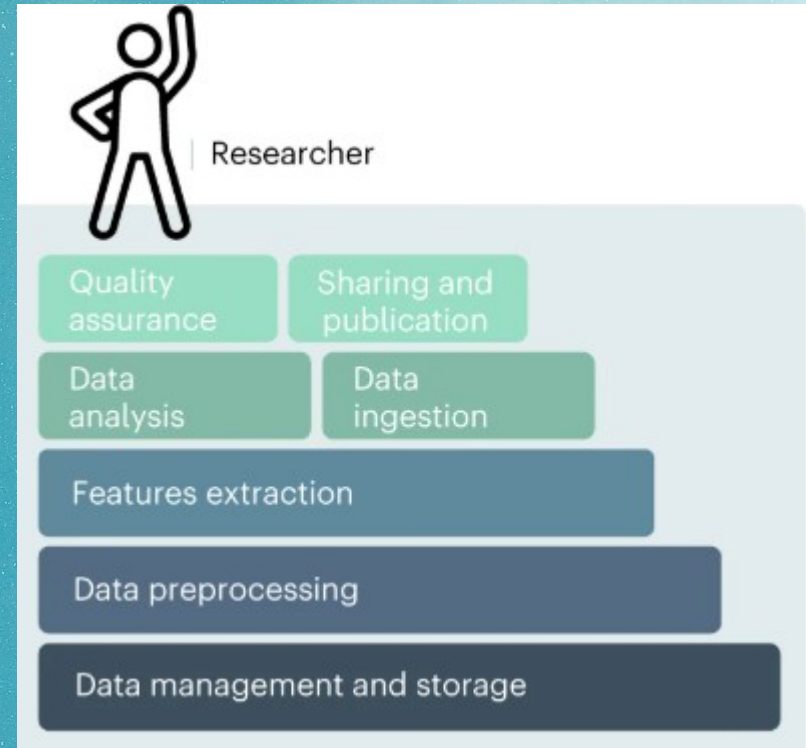
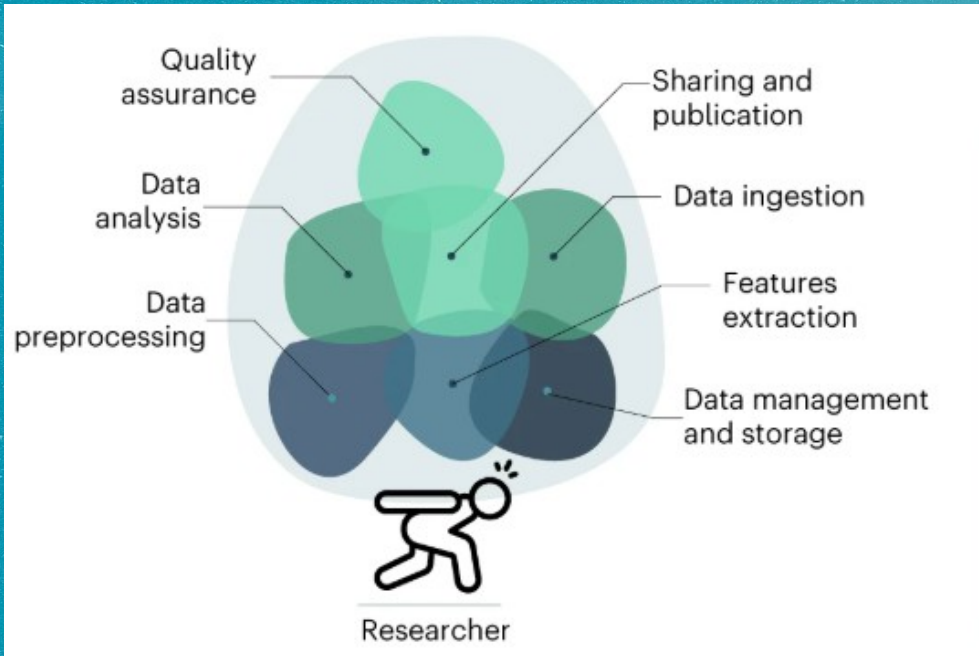
# Áttekintés

- Adattárolási és számolási platformok különböző tudományterületeken
- Adatok és számolások együttes reprodukálhatósága és megosztása
- HUN-REN Cloud és ARP
- Specifikáció és a rendszermérnök szerepe

# Neuroscience

- <https://brainlife.io/about/>
- Hayashi, S., Caron, B.A., Heinsfeld, A.S. et al. brainlife.io: a decentralized and open-source cloud platform to support neuroscience research. Nat Methods 21, 809–813 (2024). <https://doi.org/10.1038/s41592-024-02237-2>
- <https://openneuro.org/>  
A free and open platform for validating and sharing BIDS-compliant MRI, PET, MEG, EEG, and iEEG data
- Brain Imaging Data Structure (BIDS)  
Ganz, M., Poldrack, R.A. Data sharing in neuroimaging: experiences from the BIDS project. Nat. Rev. Neurosci. 24, 729–730 (2023). <https://doi.org/10.1038/s41583-023-00762-1>

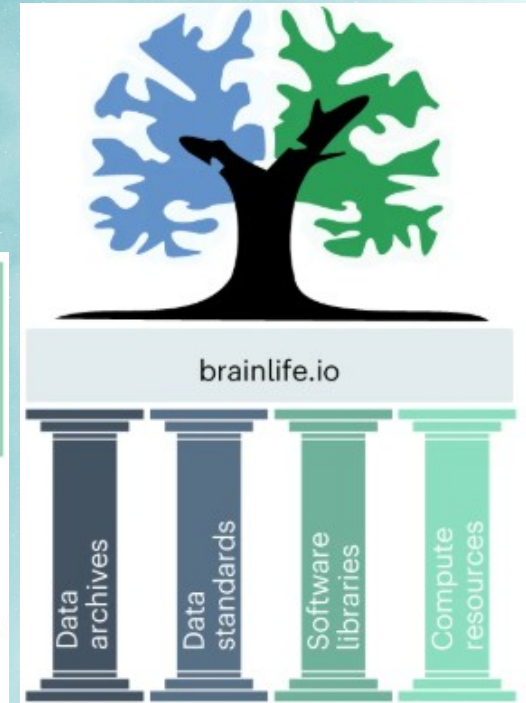
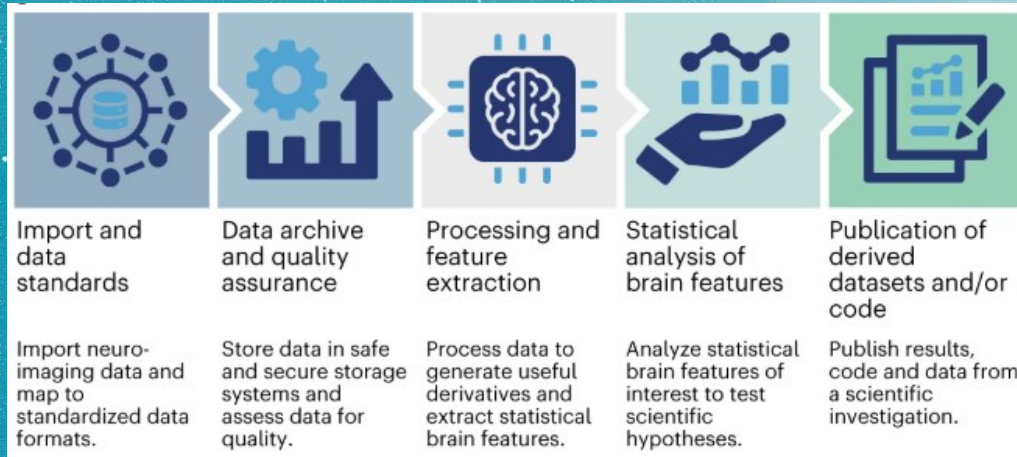
# Az integrált infrastruktúra ígérete



CREDIT: <https://doi.org/10.1038/s41592-024-02237-2> , első ábra a és b panel

# Neuroscience

- Open-source
- MRI, EEG, MEG (OpenNeuro.org archive)
- Software as a service (SaaS) megoldás
- Kulcsrakész platform HPC és Cloud erőforrásokkal



CREDIT: <https://doi.org/10.1038/s41592-024-02237-2> , első ábra c és f panel

# Genomika

- National Advisory Council for Human Genome Research (NACHGRI)  
NHGRI Analysis, Visualization, and Informatics Lab-space (AnVIL)
- [https://www.genome.gov/sites/default/files/media/files/2022-05/NACHGR\\_May2022\\_AnVIL\\_concept.pdf](https://www.genome.gov/sites/default/files/media/files/2022-05/NACHGR_May2022_AnVIL_concept.pdf)
- <https://anvilproject.org/>

# Genomika

## AnVil



### Terra

Collaborate in Terra, AnVIL's secure, scalable, cloud compute environment.

[Launch](#) [Learn More](#)



### AnVIL Data Explorer

Build cross study cohorts for analysis in Terra.

[Datasets](#) [Learn More](#)



### Dockstore

Create and share Docker-based workflows.

[Launch](#) [Learn More](#)



### NCPI

Interoperate with other NIH data commons.

[Learn More](#) [Datasets](#)



### Bioconductor

Analyze genomic data in the R statistical language.

[Learn More](#)



### Galaxy

Run batch analysis workflows and interactive visualizations.

[Learn More](#)



### Jupyter

Run interactive analysis with python or R.

[Learn More](#)



### Seqr

Identify disease-causing variants.

[Launch](#) [Learn More](#)

CREDIT: <https://anvilproject.org/> oldalról printscreen

# Genomika

## AnVil néhány modulja

- A Terra egy felhőalapú platform a bioinformatikai adatelemzéshez, amely integrálva van az AnVIL-lal. Lehetővé teszi a kutatók számára az adatok tárolását, elemzését és megosztását egy biztonságos és skálázható környezetben.
- A Dockstore egy nyílt platform Docker-alapú bioinformatikai eszközök és munkafolyamatok megosztására és keresésére. A Docker konténerek segítségével a kutatók könnyedén megoszthatják és reprodukálhatják az elemzéseiket.
- A Galaxy egy nyílt forráskódú, web alapú platform a bioinformatikai adatelemzéshez. Lehetővé teszi a felhasználók számára komplex munkafolyamatok összeállítását és futtatását, valamint az eredmények interaktív módon történő megjelenítését.
- Seqr egy genomikai elemző platform, amely a ritka betegségek genetikai okainak azonosítására összpontosít. Segítségével a kutatók és klinikusok szekvenálási adatokat tölthetnek fel, elemezhetnek és értelmezhetnek, hogy feltárják a betegségeket okozó genetikai variánsokat.




# Csillagászat

## Egy távcső a Vera C. Rubin Observatory Platformja

### Rubin Science Platform

#### Portal


Discover data in the browser



Learn more about the portal.

#### Notebooks


Process and analyze LSST data with Jupyter notebooks in the cloud

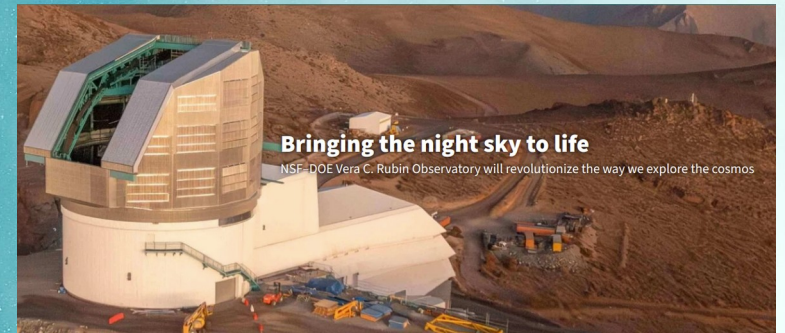


Learn more about notebooks.

#### APIs

Learn how to programatically access data with Virtual Observatory interfaces





CREDIT:  
<https://rubinobservatory.org/> oldalról  
printscreen

CREDIT: <https://data.lsst.cloud/> oldalról printscreen

# Hidrológia

## HYDROSHARE

### How can HydroShare support your research?



Sharing research products with colleagues and others using FAIR Data Principles



A dependable data management and publication solution for hydrologic data types and models



A gateway to cloud compute resources using Jupyter and Matlab



Tools for collaboration, enabling you to manage access and more easily work together



Tools for engaging students in active learning



Compliance with funding agency or journal publisher requirements for sharing data



A REST web service API and Python client for automated access



Discover and access data and models published by others



Web apps for visualizing, analyzing, and running models on data in HydroShare

# Dataverse alapú adatrepozitóriumok

- A Harvard IQSS (Institute for Quantitative Social Science) égisze alatt fejlesztett és gondozott, nyílt forráskódú Dataverse-re épülő adatrepozitóriumi rendszerek, amilyen például a Harvard Dataverse és HUN-REN Adat Repozitórium is mára elterjedtek.
- 2006 projekt indulása, ...,2010 API bevezetése,...,2017 nemzetközi adaptációk elindulása,...,2021 Kibővített adatvizualizációs eszközök, szoftvercélú tárolás támogatása (pl. Jupyter notebook-ok),..., 2024 Dataverse 6.5 verziója
- Főként intézményi (pl. HUN-REN) és multidiszciplináris repozitórium
- Centralizált megoldást kínál kutatócsoporttól és tématerülettől függetlenül
- Általános „intézményi” keret: a kutatási adatmenedzsment-szabályzatok, pályázati előírások (pl. EU-s) sokszor előírják, hogy a kutatás végeztével az adatok bizonyos formában, DOI-val ellátva, hosszú távra, nyilvánosan elérhető módon legyenek archiválva. A Dataverse könnyen beilleszthető ilyen intézményi folyamatokba.

# Domain-repozitóriumok

- Az előzőkben felsorolt tudományterületi példák archívumai:

Neuroimaging: OpenNeuro (BIDS formátummal),...

Csillagászat: CDS Strasbourg, ESO adatarchívum,...

Genomika: EMBL-EBI, NCBI archívumok

- A domain-specifikus rendszerek jobban támogatják a reprodukálhatóságot egy adott szakterületen, mert:
  - Standardizált adatformátumokat követelnek meg (pl. BIDS neuroimaging esetében, FASTQ/SRA a szekvenálásnál stb.)
  - Beépített validálási vagy elemzési pipeline-okat is kínálhatnak
  - Speciális metaadat-sémákat és követelményeket alkalmaznak, így a szakmai közönség azonnal tudja, mit várhat az adott adatállománytól.

# Áttekintés záró gondolatai

- HUN-REN Cloud lehetőséget adhat egy-egy területnek, egy-egy műszernek integrált platform létrehozásában. Nagy lépés lenne az integrált platformok megjelenése.
- Hazai számítási infrastruktúrák (pl. Szuperszámítógép, HUN-REN Cloud) integrált platformba ágyazott elérése. Külföldi példa ilyenre a [brainlife.io](https://brainlife.io)
- Kényelmes átjárhatóság biztosítása idehaza a domain specifikus repozitóriumok és dataverse típusú általános repozitórium között

# Adatgyűjtési rendszer építése

- Célok meghatározása (pl. egy integrált platform része legyen monitorozás és analízis megkönnyítése érdekében)
- Ha meg vannak a célok specifikáció készítése
- Ki irányítja a specifikáció készítését: rendszermérnök (Hol van ilyen képzés Magyarországon?)
- Specifikáció után tervezés és megvalósítás.

# Eötvös-inga Adatgyűjtő Rendszer – Specifikáció (Mockup)

## 1. Cél és Hatókör

- **Fő cél:** Biztosítani, hogy az Eötvös-inga mérésének minden releváns paramétere (torziós szög, hőmérséklet, páratartalom) folyamatosan és megbízhatóan elérhető legyen a kutatók és üzemeltetők számára.
- **Hatókör:** Tartalmazza a rendszertervezést, a szenzorok kiválasztását, a kommunikációt és adatkezelést, valamint az üzemeltetők számára szükséges dokumentációt.

## 2. Fizikai Mérőberendezés – Eötvös-inga

- Egy torziós ingarendszer, amelyet forgó platformra erősítünk.
- A torziós szál anyaga: foszforsavval edzett volfrám (példa).
- A skálaleolvasás részlegesen digitalizált (kamera + szoftveres képfeldolgozás).

## 3. Környezeti Paraméterek Mérése

- **Hőmérséklet**
  - Digitális szenzor ( $\pm 0,1$  °C pontosság).
  - Minta: DS18B20 vagy RTD (PT100/PT1000).
- **Páratartalom**
  - Kapacitív páratartalom-érzékelő ( $\pm 3\%$  RH pontosság).
- **Mintavételezési gyakoriság**
  - 30 másodpercenkénti leolvasás, figyelembe véve a szenzorok válaszidejét.

# Eötvös-inga Adatgyűjtő Rendszer – Specifikáció (Mockup)

## 4. Adatátviteli és -gyűjtési Megoldás

- **Mikrokontroller/SBC:** A szenzoradatokat gyűjtő központi egység (pl. Raspberry Pi).
- **Protokoll:** I2C a hőmérséklet- és páraszenzorhoz, SPI (vagy dedikált) interfész az optikai enkóderhez.
- **Hálózat:** Helyi Ethernet/Wi-Fi a laborban, MQTT üzenetszolgáltatóval.

## 5. Adatkezelés és Tárolás

- **Lokális rögzítés:** MicroSD vagy SSD a mikrokontrolleren, legalább 1 hét kapacitással.
- **Távoli adatbázis:** Időalapú (Time-Series) DB (InfluxDB, Prometheus), napi mentéssel.
- **Biztonsági mentés:** Heti mentés NAS-ra, opcionálisan felhőbe.

## 6. Rendszerintegráció és Régi Eszközök

- A torziós inga meglévő optikai skálájának képi leolvasása (kamera).
- Képelemző szoftver, amely valós időben konvertálja a skálaállást digitális értéké (szög).
- Időbélyeg-szinkron a mikrokontroller belső órájával (NTP).

## 7. Kalibráció és Karbantartás

- **Hőmérséklet-érzékelő:** Éves kalibrálás külső laborban, referencia-hőmérővel.
- **Páratartalom-érzékelő:** Féléves ellenőrzés sóoldatos (33% RH, 75% RH) referencia-pontokon.
- **Optikai enkóder és kamera:** Mechanikai ellenőrzés (felfüggesztés, szálfeszesség) minden karbantartási ciklusban.



# Eötvös-inga Adatgyűjtő Rendszer – Specifikáció (Mockup)

## 8. Riportolás és Vizualizáció

- **Real-time kijelzés:** A labor számítógépén futó dashboard (pl. Grafana).
- **Riasztások:** E-mail értesítés, ha a hőmérséklet  $\pm 0,2$  °C eltérést mutat a beállított küszöbtől.

## 9. Üzemeltetők Feladatai és Felelősségi Körök

- **Napi Ellenőrzés:** Gyors állapot- és adatellenőrzés, esetleges hibaesemények rögzítése.
- **Havi Karbantartás:** Szenzorburkolatok tisztítása, kamera élességellenőrzése.
- **Rendszermérnök:** Felelős a hardver/szoftver frissítésekért, kalibráció szervezéséért, dokumentáció naprakészen tartásáért.

## 10. Dokumentáció és Verziókezelés

- **Specifikáció:** Ezt a mockup anyagot bővítve kell karbantartani (aláírások, verziók, változási napló).
- **Felhasználói kézikönyv:** Tartalmazza a szenzorok csatlakoztatásának lépéseit, a hibaelhárítási útmutatót és a biztonsági óvintézkedéseket.
- **Kísérleti Napló:** A mindennapi mérések, események (pl. különleges kísérleti konfigurációk, rendellenességek) rögzítésére.

# Rendszermérnök szerepe és felelőssége az Eötvös-inga méréseknél (Mockup)

- **Rendszerszintű tervezés**

Meghatározza a mérőberendezés (Eötvös-inga), a környezeti szenzorok és az adatgyűjtő infrastruktúra közötti teljes architektúrát, beleértve a kommunikációs protokollokat és a szoftverkörnyezetet.

- **Követelmények egyeztetése**

Együttműködik a kutatókkal és az üzemeltetőkkel, hogy pontosan felmérje a mérés célját, a szükséges pontosságot, a környezeti feltételeket, valamint az üzemeltetés gyakorlati igényeit (pl. karbantartási ciklusok).

- **Technológiai megoldások kiválasztása**

Kiválasztja a megfelelő szenzorokat (hőmérséklet, páratartalom stb.), adatgyűjtő rendszereket (mikrokontroller, SBC), és meghatározza az optimális mintavételezési ütemet, hogy a rendszer a lehető legpontosabb adatokat szolgáltatassa.

- **Integráció és beüzemelés**

Gondoskodik a meglévő vagy régi komponensek (például részlegesen digitalizált skálaleolvasás) zökkenőmentes összekapcsolásáról a modern digitális eszközökkel, biztosítva az időbélyegek és a kalibrációk pontos kezelést.

# Rendszermérnök szerepe és felelőssége az Eötvös-inga méréseknél (Mockup)

- **Adatbiztonság és -kezelés**

Kidolgozza a helyi és távoli adattárolási stratégiát, a hibabiztos adatátvitelt (pl. pufferek, redundáns mentés), valamint a hozzáférési jogosultságokat a kutatócsoport számára.

- **Dokumentáció és karbantartás**

Részletes leírásokat, telepítési útmutatókat, műszaki rajzokat készít, valamint nyomon követi a rendszerverziókat. Felügyeli a rendszeres kalibrációt és a hosszú távú karbantartási terveket (pl. szenzorcsere, firmware-frissítések).

- **Minőségbiztosítás és hibaelhárítás**

Felelős a rendszer teljesítményének folyamatos ellenőrzéséért, a hibajelenségek gyors detektálásáért és megoldásáért. Szükség esetén javaslatot tesz korszerűsítésre (pl. pontosabb szenzorok, hatékonyabb adatfeldolgozás).

- **Kommunikáció és oktatás**

Kapcsolatot tart a többi szakemberrel (kutatók, műszaki személyzet, informatikusok) és segíti az üzemeltetők munkáját. Képzési anyagokat biztosít, konzultációkat tart, hogy mindenki értse a rendszer működését és lehetőségeit.

# Záró Gondolatok

- Specifikáció, mint az első lépés a tervezés és megvalósítás felé
- A rendszermérnök, mint koordinátor összefogja a csapat munkáját
- A sikeres megvalósítás csapatmunka eredménye
- Perspektíva integrált platformok a kutatási infrastruktúráinknak

**Köszönöm!**