

# Adattár (repozitórium) használatot elősegítő eszközök fejlesztése OpenBioMaps - biológiai adatkezelő – platformon

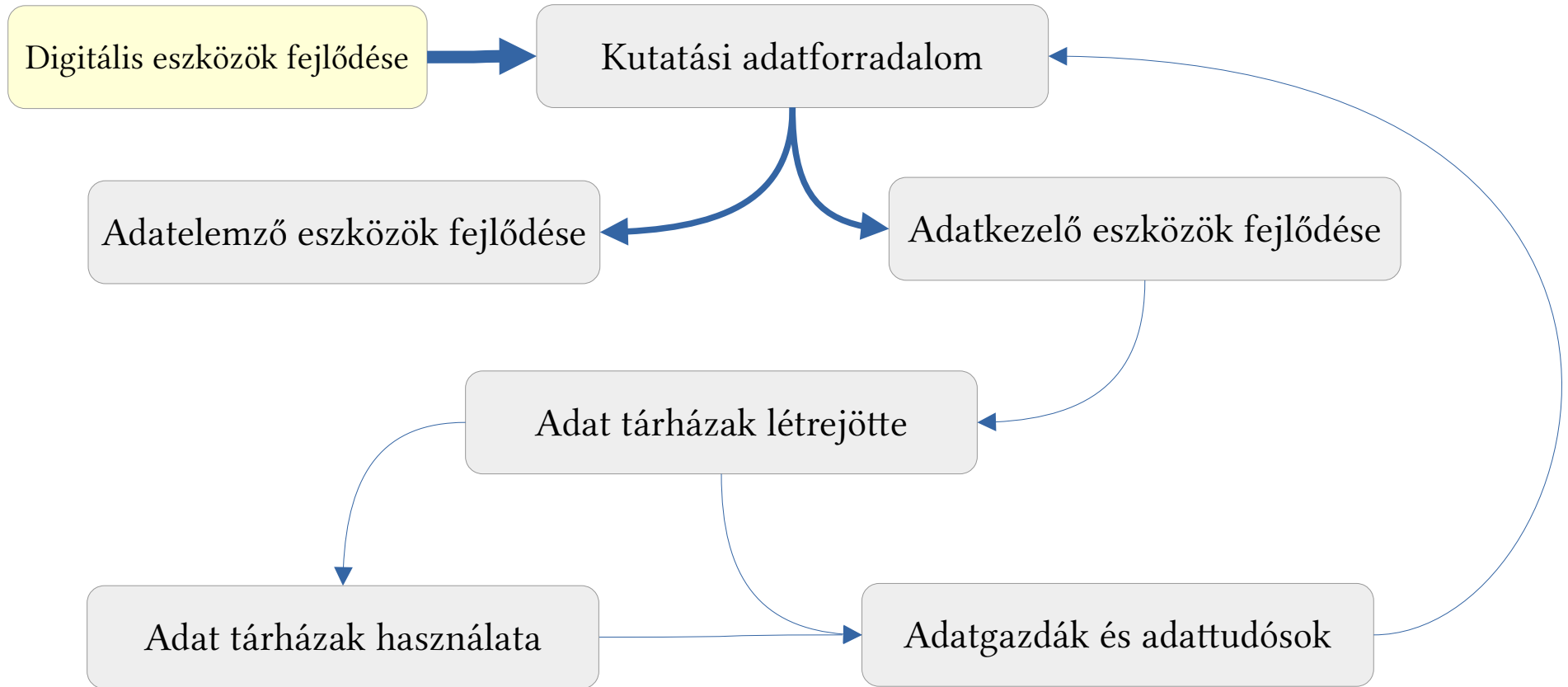


Dr. Bán Miklós

2022 január 18.

Debreceni Egyetem, MTA-DE Viselkedésökológiai Kutatócsoport

# Tudományos adattárak és adatgazdászok



# Mi az OpenBioMaps?

Természetvédelmi és biodiverzitás kutatási adatkezelő platform

## Digitális adatstruktúrák létrehozása:

- Kollaboratív munkaterületek alapítása
  - Szerverek létrehozása
  - Adatbázisok alapítása

## Digitális adatok létrehozása:

- Megfigyelési adatok adatbázisba juttatása
- Már digitalizált adatbázisok behozatala



Adatok transzformálása,  
kiegészítése és elemzése

Adatok keresése és  
megjelenítése

Adatok továbbjuttatása  
más eszközök felé

# Biológus tudományos munkafolyamat

- Mikor kerül a képbe a repozitórium?
  - Amikor az adat keletkezik?
    - Nem jellemző, ez inkább az adatbázisok területe, mint pl. az OBM
  - Amikor elemezhető állapotba kerül?
    - Ez jó lenne, de legfeljebb GIT repozitóriumokba kerül be az ilyen adat
  - Amikor elkészült az elemzés?
    - Általában, de ezek egy része kerül supplementary materials-ba is





# Milyen repo eszközöket fejlesztettünk?

## 1) Általános repozitórium kezelő eszköz

- 1.1) Szerver szinten és projekt szinten is lehet repó kapcsolatokat regisztrálni
- 1.2) Tetszőleges számú és típusú repozitórium kapcsolat regisztrálható
- 1.3) Tetszőleges repozitórium attribútumok definiálhatóak

## 2) Repozitórium használat API-k közötti kapcsolat használatával

- 2.1) Központi autentikáció egyéni autentikáció helyett: OBM alkalmazás API kommunikál a repozitórium API-val, projektenként mindenki egy bejelentkezést használ, nincsenek egyéni repo felhasználók alapértelmezetten.

## 3) Repozitórium opciók célirányos egyszerűsítése

## 4) Dataverse





# Milyen repo eszközöket fejlesztettünk?

## 1) API tetszőleges adattárakhoz

1) Átlátszó funkciók adattár specifikus funkciók felett

## 2) API a Dataverse adattár számára

- Adatgyűjtők (dataset) listázása (OBM)
- Adatgyűjtő leíró adatainak lekérdezése (transparent: dataverse)
- Gyakori tulajdonságok lekérdezése adatgyűjtőből (OBM)
- Adatgyűjtő publikálása (transparent: dataverse)
- Új adatgyűjtő létrehozása (OBM)
- Fájl létrehozása (transparent: dataverse)
- Fájl lecserélése (transparent: dataverse)
- Fájl eldobása (transparent: swordv2)
- Adattárak keresése (OBM)





# Milyen repo eszközöket fejlesztettünk?

## 1) R csomag

- Tetszőleges adattár kapcsolat kiválasztása
- Adatgyűjtők létrehozása
- Fájlok feltöltése adatgyűjtőkbe
- Fájlok letöltése adatgyűjtőkből
- Fájlok törlése adatgyűjtőkből
- R környezet mentése adatgyűjtőbe





# Milyen repo eszközöket fejlesztettünk?

## 1) Webes alkalmazás

- Elérhető adattárak listázása
- Adattár tulajdonságok lekérdezése
- OBM-ben eltárol (belső tartós azonosítóval ellátott) lekérdezések adatgyűjtővé alakítása
- Adatgyűjtő kiegészítése
- Adatgyűjtő publikálása





# Mit lehetne vajon még tenni az ügy érdekében?

- Figyelembe kell venni a kutatói munkafolyamatot és meg kell keresni azokat a pontokat, ahol tudunk valami valós segítséget nyújtani a felhasználóknak anélkül, hogy tőlük a jelenlegi szokásaik lecserélését várnánk.
- Az elemzési folyamat során is előkerülhetnének az adat táruk mint a közösségi munka színtere és elsődleges archiválási terület.
  - 1) Ehhez távoli fájlrendszerként tudnia kellene működni az adat tárnak
  - 2) Beépített GIT/rsync funkciók kellene támogatnia
  - 3) Felhő tárolási módszerrel egyénileg klaszterezhetőnek kellene lennie:  
azaz támogatnia kellene a repozitórium csak interface működési módot

