

Excel-tábláktól a szabványos digitális archívumig

Zagyva Natália



HAGYOMÁNYOK
HÁZA

XI. Adatgazdász-találkozó
Hagyományok Háza, 2026.09.02.



 FOLKLÓRADATBÁZIS

GYŰJTŐK

HELYEK

TÉRKÉP

CIKKEK

RÉSZLETES KERESŐ

Bejelentkezés 

Ezt keresem...

KERESÉS

CIKKEK

Az első a zenekaré?

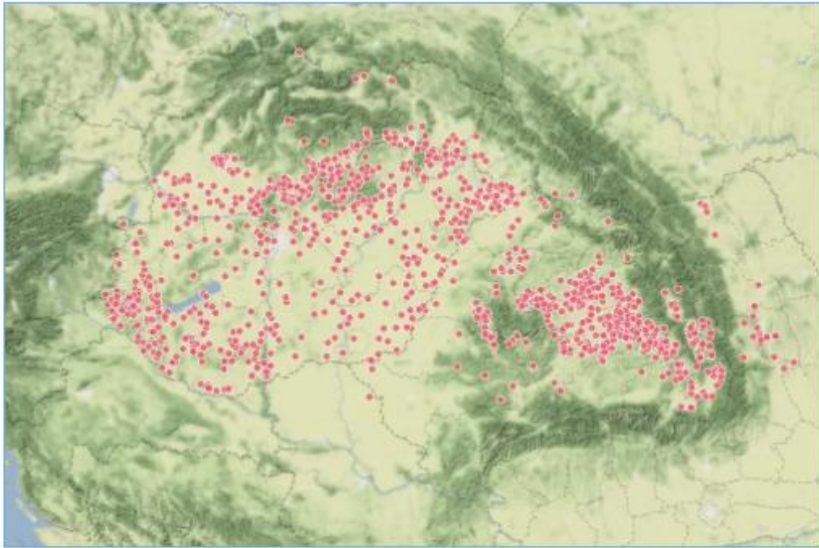
NÉMETH LÁSZLÓ • 2026. AUGUSZTUS 25.



A magyar népi tánczene dallamainak jelentős részét a falusi közösségekben tánc közben is énekelték. Számos tájegységen a mulatónóták zenekari kíséretes éneklésének is jelentős szerepe volt. De hogyan történt a gyakorlatban egy-egy nóta kezdése, a dallamok váltása?

TOVÁBB→

TÉRKÉP



A GYŰJTEMÉNYBŐL



1994.03.02. Budapest
Halmos Béla, Németh István,
Vavríneckz András

243

1936.04. Sárpilis
Lajtha László

<https://folkloradatbazis.hu>

A Folklór Archívum részletesen feldolgozott anyagaiból online elérhető:

- 7103 folklórgyűjtés
- 1505 óra videó
- 5139 óra hang

0:01 / 2:31

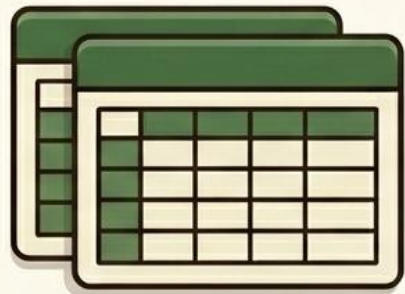


CIKLUS	1. ciklus		2. ciklus							
TÉTEL	szökös	sűrű csárdás	lassú csárdás				sűrű csárdás		meny...	menyasszonykísé
DALLAM	1. dallam	2. dallam	3. dallam	4. dallam	5. dallam	6. dallam	7. dallam	8. dallam	9. dal...	10. dallam
TÁNCSEKASZ	szökös	sűrű csárdás	lassú csárdás				sűrű csár...	sűrű csárdás		
FOLKLÓRSZÖVEG	c.. lakod									

A Folklór Archívum gyűjteménykezelésének megújítása



Megoldandó problémák



Excel-alapú nyilvántartás

Töredezett, redundáns, nehezen kereshető nyilvántartás, nem szabványos metaadat-szerkezet. (több ezer Excel-tábla)



Kézi adatmozgatás

A digitális reprezentációk és használati fájlok létrehozása manuálisan, verziókezelés nélkül történik. A *born digital* anyagok beadási folyamata nem automatizált.



Offline fizikai hordozók

Megőrzés külső adattárolókon (HDD, LTO), hálózati elérés nélkül a sérülés és elavulás nehezen detektálható, az adatvesztés észrevétlen maradhat.



Elszigetelt hozzáférés

Csak a Folklóradatbázisban részletesen feldolgozott és publikált tartalmak érhetők el online. A gyűjtemény nagy része csak helyben, a könyvtárban kutatható. Nincs API, nincs gépi adatcsere-lehetőség.

A Folklór Archívum gyűjteménykezelésének megújítása



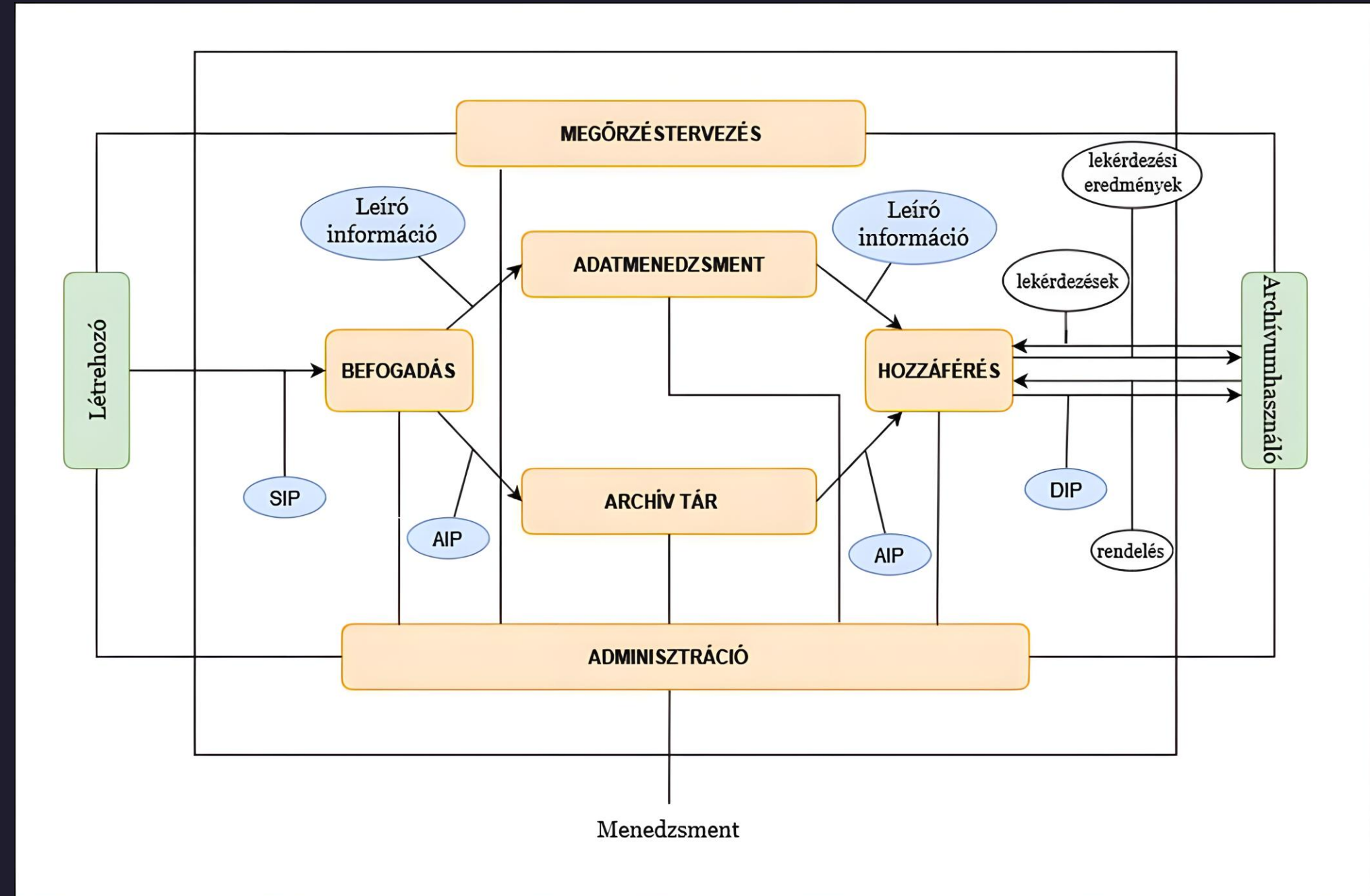
Fenntartható és jövőálló megoldás:

Szabványokon alapuló gyűjteménykezelés és
hosszú távú megőrzés
nyílt forráskódú szoftverek alkalmazásával

Elméleti keretrendszer:

OAIS referenciamodell

A modell meghatározza az archívum
környezetét, funkcionális egységeit és az
adatok életciklusát kísérő információs
csomagokat.



OAIS – Nyílt archívumi információs rendszer

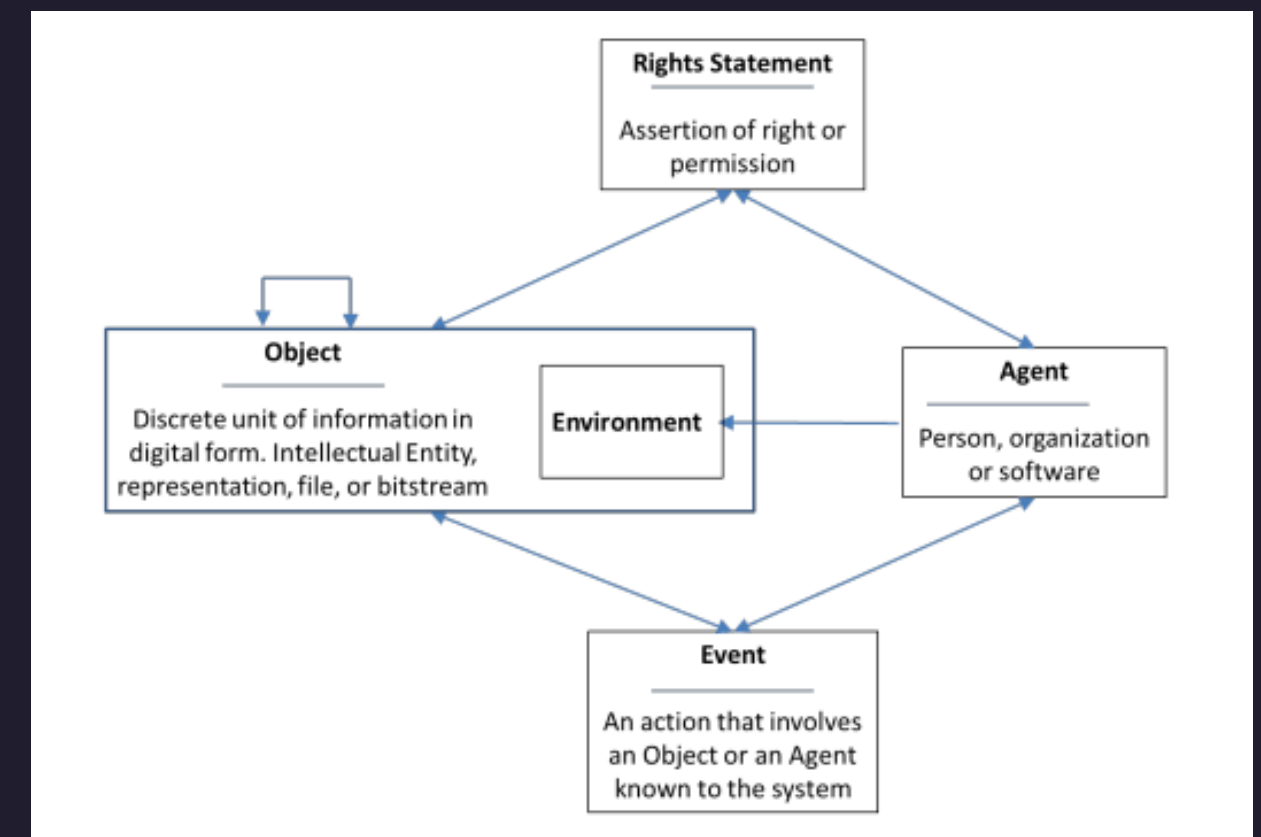
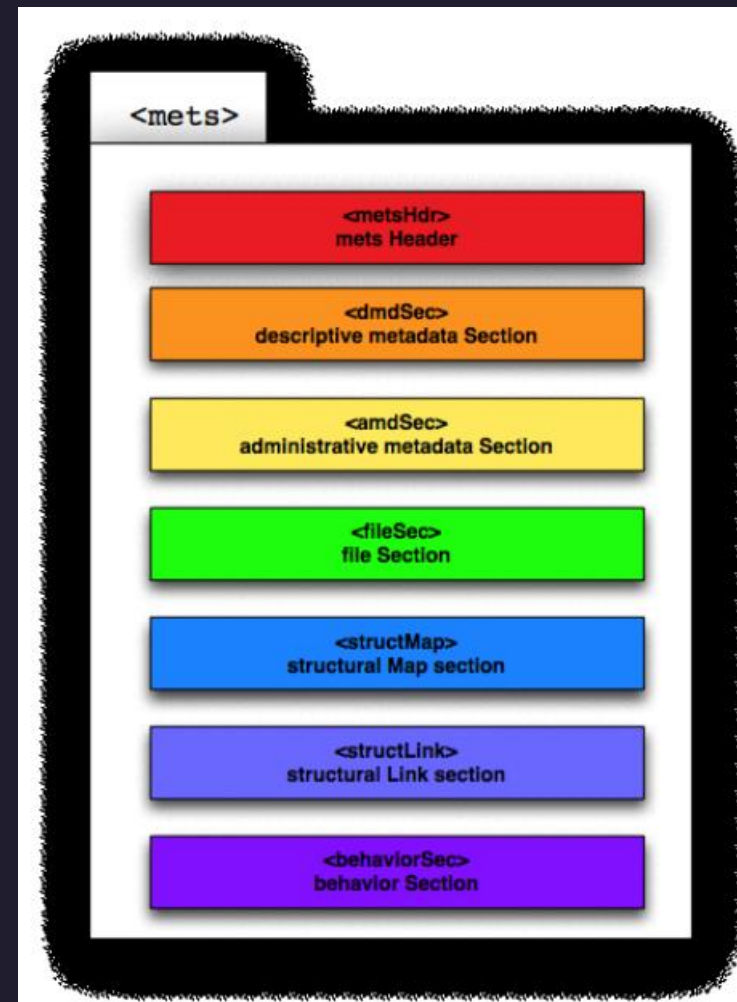
Ilácsa Szabina és Bódog András (KöSZI) ábrája az MSZ ISO 14721:2022 szabvány alapján (Bódog, 2022, p. 23)

A Folklór Archívum gyűjteménykezelésének megújítása



Metaadat-szabványok

- **Strukturális:**
METS (XML konténer)
- **Megőrzési:**
PREMIS (integritás, eseménynaplózás, jogosultságok)
- **Leíró:**
ISAD(G) – hierarchikus levéltári leírás, EAD (XML séma a gépi adatcseréhez),
Records in Contexts (gráfalapú levéltári leírás),
Dublin Core (az interoperabilitáshoz)



Nemzetközi kutatások és tapasztalatok

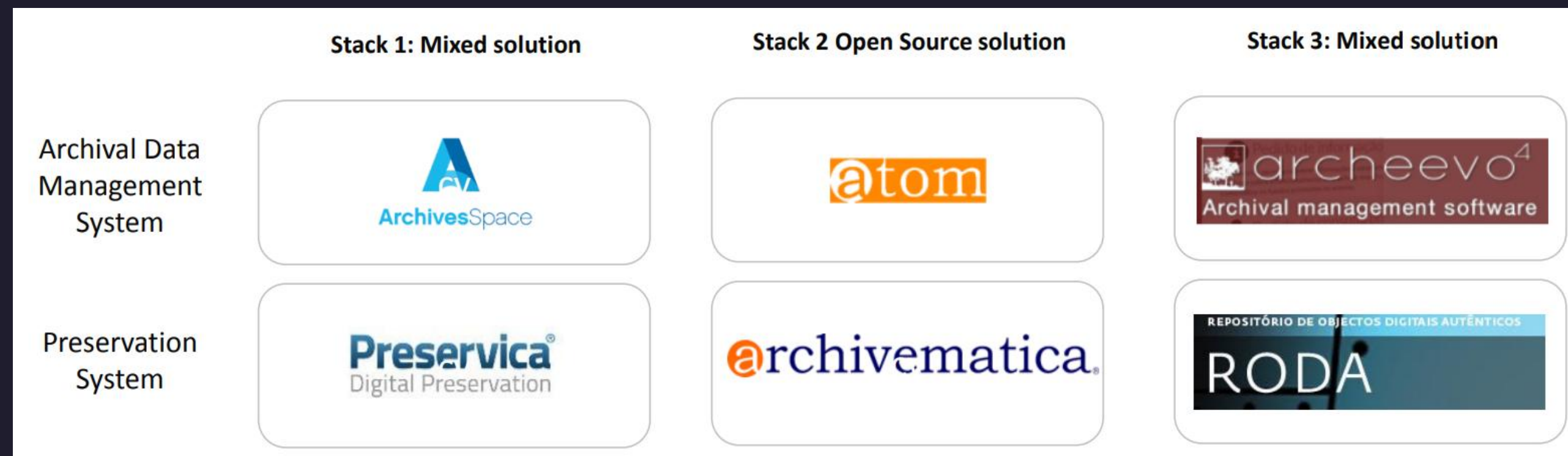


Összegyűjtöttük a nemzetközi ajánlásokat és jógyakorlatokat.

ISA² program (2016–2020): kutatás az európai közszolgáltatások interoperabilitásának támogatásáról.

2018-as jelentés az addigi eredményekről: „Standard-Based Archival Data Management, Exchange and Publication” (315 oldal)

A tanulmány háromféle szoftvercsomagot ajánl egy konkrét intézmény gyűjteménykezelő és megőrzési rendszerének megvalósítására:



„Software stack” (Segura et al., 2018, p. 89)

Gyűjteménykezelő rendszerek



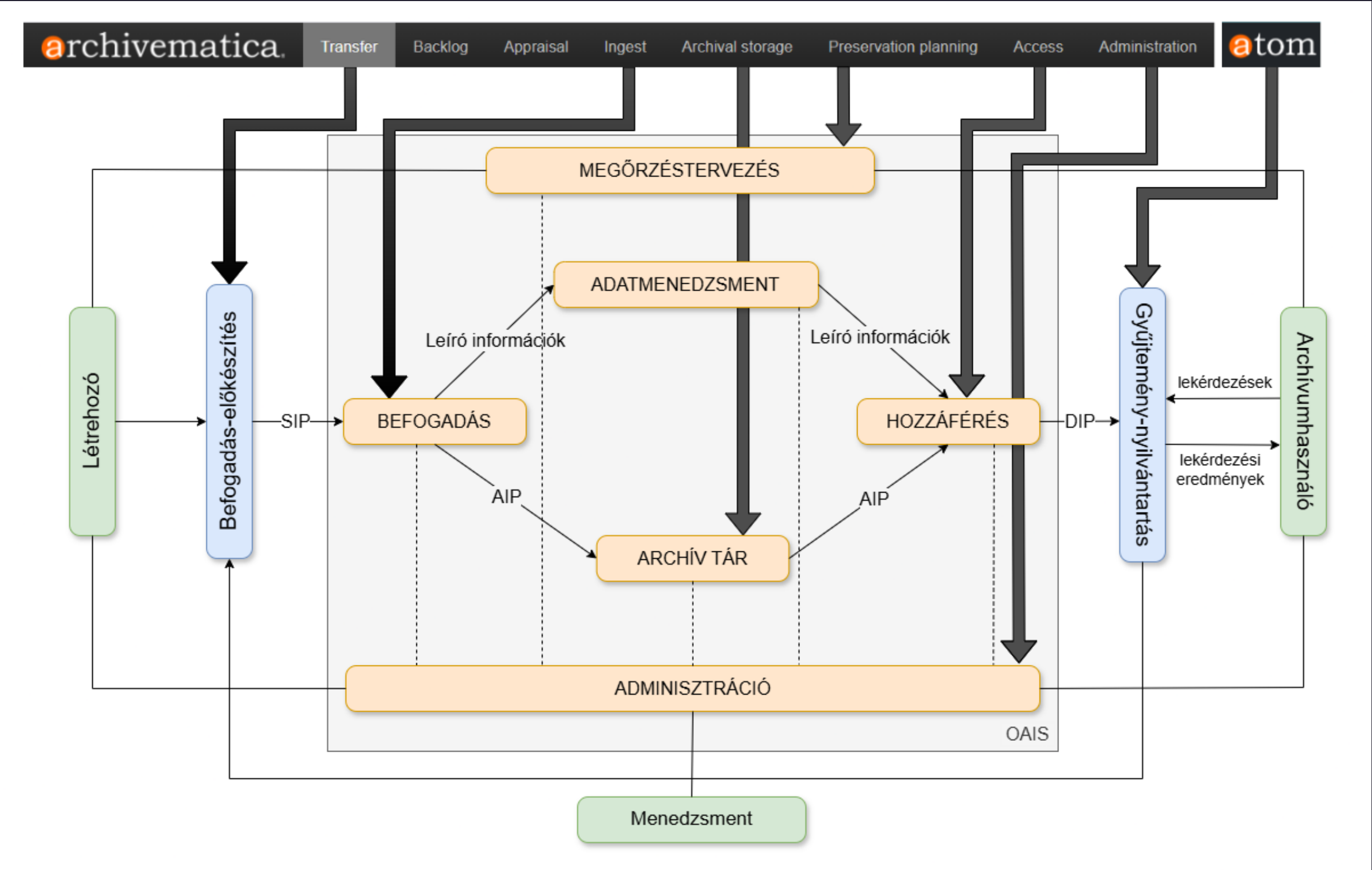
	ArchivesSpace	AtoM
Alapértelmezett leírási metaadat-szabvány	DACS – kompatibilis az ISAD(G) szabvánnyal (Society of American Archivists, 2013)	ISAD(G)
Nyelvi támogatás	angol, de kiegészíthető saját nyelvi verzióval	többnyelvű, van magyar nyelvi változata
Elterjedtség	leginkább az Egyesült Államokban használják, de vannak európai példák is	nemzetközi, van néhány magyar példa is
Fizikai objektumok tárolási helyének kezelése	robosztus	korlátozott
Digitális objektumok kezelése	a fájlokat más rendszerbe kell betölteni, itt csak az elérési útvonal adható meg	a fájlok betölthetők a rendszerbe
Támogatás	széles, főként amerikai közösségi támogatás, tagdíj-alapú hozzáférés bizonyos szolgáltatásokhoz	széles nemzetközi közösségi támogatás, ingyenes hozzáférés minden támogatói anyaghoz

Hosszútávú megőrzésre szolgáló rendszerek



	Archivematica	RODA CE
Irányadó metaadat-szabvány	METS, PREMIS (főként technikai és megőrzési fókusz)	METS, PREMIS, EAD, Dublin Core (erős leíró fókusz)
Nyelvi támogatás	angol, szükség esetén kiegészíthető saját nyelvi verzióval	többnyelvű, pl. angol, portugál, de részben elkészült már a magyar nyelvi változat is
Elterjedtség	globális	elsősorban Európa, különösen Portugália
Alapfilozófia	munkafolyamat-motor (pipeline): a fájlok feldolgozására és csomagolására fókuszál.	teljes körű repozitórium: a bekerüléstől a hosszú távú tárolásig és menedzsmentig mindent egyben kezel
Adattárolás	nincs saját beépített tárhelye, a kész AIP-okat külső tárolóra küldi	saját beépített tárhely
Normalizálás	rendkívül erős: a Format Policy Registry segítségével automatikusan konvertál	támogatott
Technológia	Python, Django, Elasticsearch, microservices architektúra	JAVA, Solr, Fedora-alapú architektúra
Integrálhatóság	kiváló: natív kapcsolata van az AtoM és ArchiveSpace rendszerekkel	jó: REST API-kon keresztül integrálható, de alapvetően zártabb, önállóbb egység
Skálázhatóság	jó: Vízszintesen skálázható extra feldolgozó csomópontok hozzáadásával. Extrém nagy fájl szám esetén a kezelőfelület és az adatbázis lassulhat.	kiváló: Alapjaiban nagyvállalati terhelésre tervezve. A konténerizált architektúra révén rugalmasan kezeli a masszív adatfolyamokat és indexelést.

Archivematica–AtoM-ökoszisztéma



archivematica | Transfer | Backlog | Appraisal | Ingest | Archival storage | Preservation planning | Access | Administration | zagyva.natalia | Connected

Transfer type: Standard | Transfer name: | Accession no.: | Access system ID: | Browse | Start transfer | Approve automatically

Transfer	UUID	Transfer start time
HH_BD_20250403_1	d0a07f6e-b16a-4c7e-9f08-f31dfe3d5fe	2026-07-13 22:20
HH_DB_20250203	Scff7001-2533-4be0-bb5a-e1ebc6cbde21	2026-06-14 17:07

HH Gyűjteményi Adatbázis | Böngészés | Keresés | Bejelentkezés

tétel 001 - 1. kazetta: Csíkszereda, 1980.06.20

Hierarchikus struktúra elrejtése

- Pávai István magángyűjteménye
 - Magnókazetta-gyűjtemény
 - 1. kazetta: Csíkszereda, 1980.06.20
 - PL_Kz_001_c
 - PL_Kz_001a
 - PL_Kz_001a.j
 - PL_Kz_001b.j2

Gyorskeresés | Keresés a hierarchikus struktúrában

Pávai István magángyűjteménye > Magnókazetta-gyűjtemény > 1. kazetta: Csíkszereda, 1980.06.20

Galéria

Alapadatok

Vágólap | Új

Főlfedezés/kutatás | Jelentések | Leltár | Böngészés listaként

Exportálás | Dublin Core 1.1 XML | EAD 2002 XML

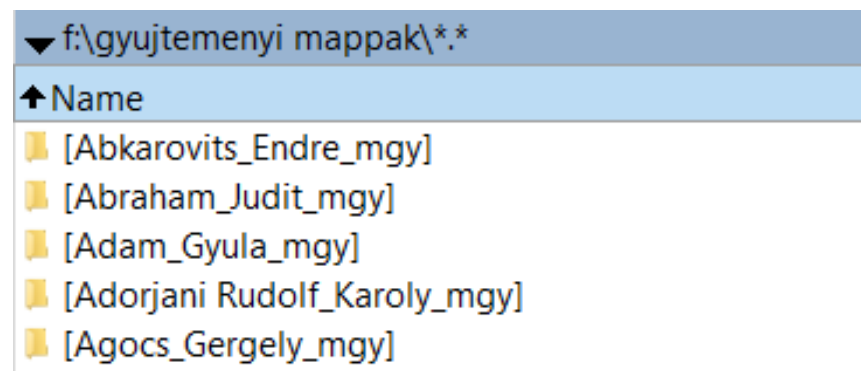
Kapcsolódó személyek és szervezetek | Lőrinc Györgyné Hodorog Luca (Tárgy)

Kapcsolódó helyek | Klézse



1. ALAPIMPORT

- Cél: A teljes gyűjtemény azonnali lefedése egy strukturált, de minimális leíró adatsorral az AtoM-ban.
- Szint: Fond, állag és sorozat szintű leírások betöltése az ISAD(G) szabványnak megfelelően.
- Megvalósítás: Digitális médiafájlok nélkül, csak a strukturális hierarchia felépítése a jelenlegi könyvtárstruktúra alapján.



2. ADATGAZDAGÍTÁS

- Cél: A leírási szintek mélyítése, a digitális objektumok biztonságos csatolása és ellenőrzött metaadatok rögzítése adatvesztés nélkül.
- Folyamat: Gyűjteményi egységenként haladva, manuális és szakértői validálással tisztítjuk a metaadatokat, az Archivemata segítségével elkészülnek az AIP-ok és az AtoM-ba kerülő DIP-ek.
- Eredmény: A tényleges médiaállományok (audio, video, fotó, stb.), jogi és adminisztratív dokumentumok (pl. adományozási szerződések) hozzákapcsolása a rekordokhoz, részletes és ellenőrzött metaadatokkal.

ArchívKapu: born digital anyagok beadási felülete



ArchívKapu – a Folklór Archívum beadási felülete

Új feltöltésElőzményekArchívum admin

1 Mappa2 Általános adatok3 Fájlok4 Beküldés

Mappa kiválasztása

Válaszd ki azt a mappát, amelynek tartalmát be szeretnéd adni az archívumba. A mappa összes támogatott médiafájlja feltöltésre kerül.



Kattints ide, vagy húzd ide a mappát

JPG · PNG · TIFF · HEIC · RAW · MP4 · MOV · AVI · MPG · MXF · MP3 · WAV · FLAC · PDF · TXT

A mappa kiválasztásakor a böngésző esetleg rákérdez, hogy beolvashatja-e a mappa tartalmát – ez még **nem** tényleges feltöltés, csak engedélykérés, hogy a böngésző lássa, milyen fájlok vannak benne. A fájlok ténylegesen csak a folyamat legvégén, a "Beküldés" gombra kattintva kerülnek fel a szerverre.

Tovább →

- Saját fejlesztésű Flask-alapú webalkalmazás
- Egyszerű, magyar nyelvű felület egy mappa feltöltésére és a mappa- és fájlszintű Dublin Core metaadatok megadására
- Az alkalmazás automatikusan létrehozza az Archivemata-kompatibilis szerkezetet:

```
20260820_083829_20210721_Karamurat_Dobrudzsa/  
├── objects  
│   ├── C0005.MP4  
│   ├── C0005T01.JPG  
│   ├── DSC00003.JPG  
│   └── TASCAM0009534.wav  
└── metadata  
    └── metadata.csv
```

- EXIF adatok vizsgálata után az alkalmazás a gyűjteménykezelő döntéseinek megfelelően átnevezi a fájlokat. A metadata.csv-ben tároljuk az eredeti fájlszerkezetet és fájln neveket.

```
20260820_083829_20210721_Karamurat_Dobrudzsa/  
├── objects  
│   ├── AG_BD_20210721_1_112325.mp4  
│   ├── AG_BD_20210721_1_112325.jpg  
│   ├── AG_BD_20210721_1_194728.jpg  
│   └── AG_BD_20210721_1_001.wav  
└── metadata  
    └── metadata.csv
```


Modernizálási lehetőség: AtoM Heratio Framework

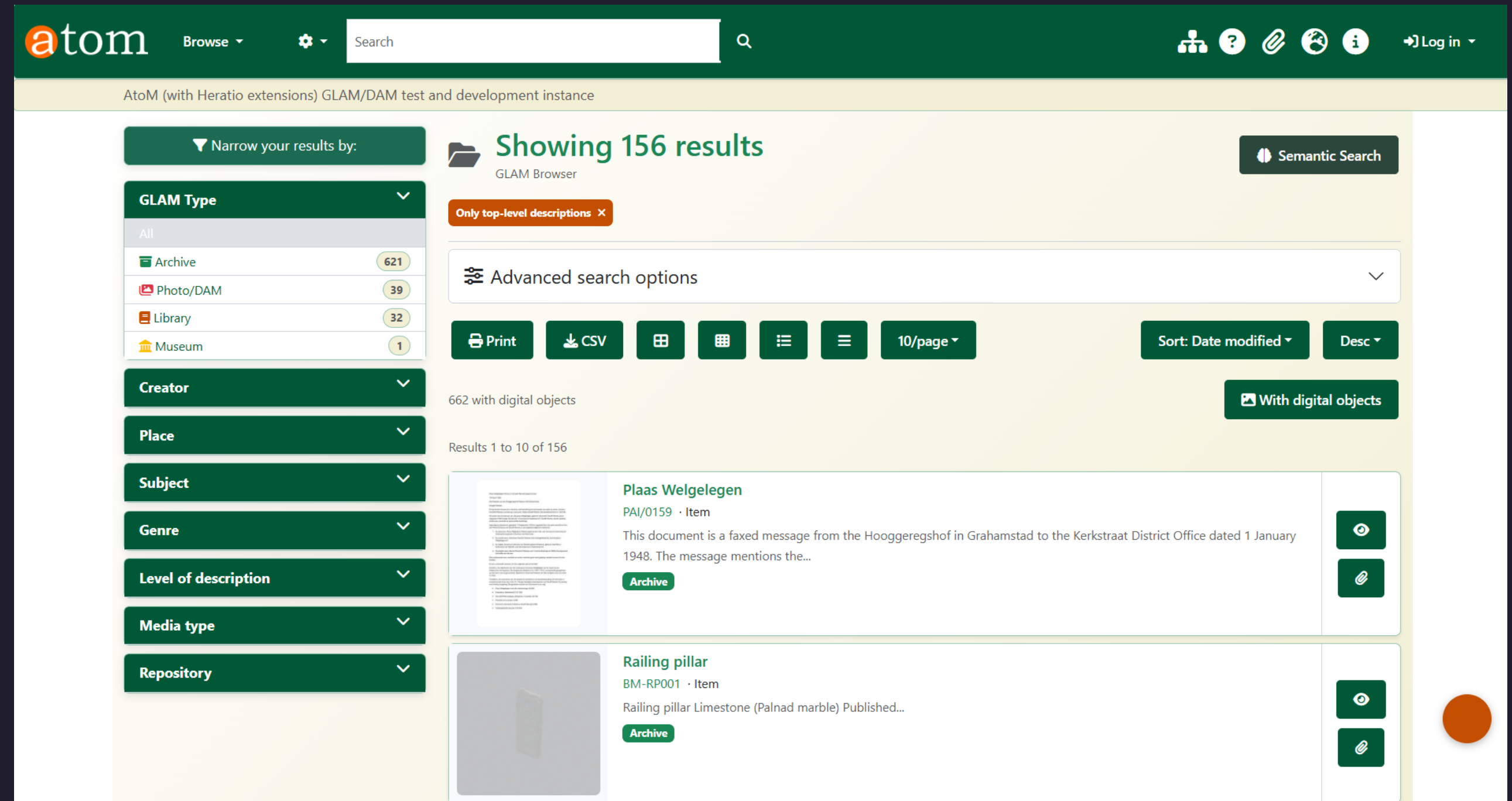


HERATIO FRAMEWORK

- AtoM-ra épülő Laravel-alapú nyílt forráskódú keretrendszer
- A teljes GLAM szektort célozza
- 62 választható bővítmény

AHG PLUGINS

- Egyesével telepíthető bővítmények az AtoM-hoz
- Jelenleg 8 bővítmény elérhető



AtoM Heratio Framework / AHG Plugins



- Több szektoros támogatás (GLAM/DAM)

Levéltárak – ISAD(G), RAD, DACS, Dublin Core

Könyvtárak – MARC-alapú bibliográfiai rekordok ISBN-kereséssel

Múzeumok – CCO (Cataloging Cultural Objects) és Spectrum 5.0 eljárások

Galériák – Kiállítások és műtárgyak katalogizálása

DAM – IPTC/XMP metaadatok digitális objektumokhoz és fényképekhez

- **AI eszközök**

Entitások kinyerése (NER) – Személyek, helyek, szervezetek automatikus azonosítása leírásokból

Összefoglaló generálása – Hosszabb leírások tömör kivonatának készítése

Fordítás – Fordítás levéltári terminológiára finomhangolva

Ellenőrző irányítópult – AI által generált javaslatok kezelése és jóváhagyása

- Digitálisobjektum-megjelenítők

Mirador – IIIF-kompatibilis képfeldolgozó megjelenítő annotációkkal

OpenSeadragon – Mély zoom nagy felbontású képekhez

Többoldalas TIFF/PDF megjelenítő – Oldalakra bontott dokumentumok böngészése

Videó-/hangrészletek – Részletek kivágása, mentése és letöltése AV-tartalomból

3D modell megjelenítő – Google Model Viewer AR-támogatással, hotspotokkal és IIIF 3D manifestekkel

AtoM Heratio Framework / AHG Plugins



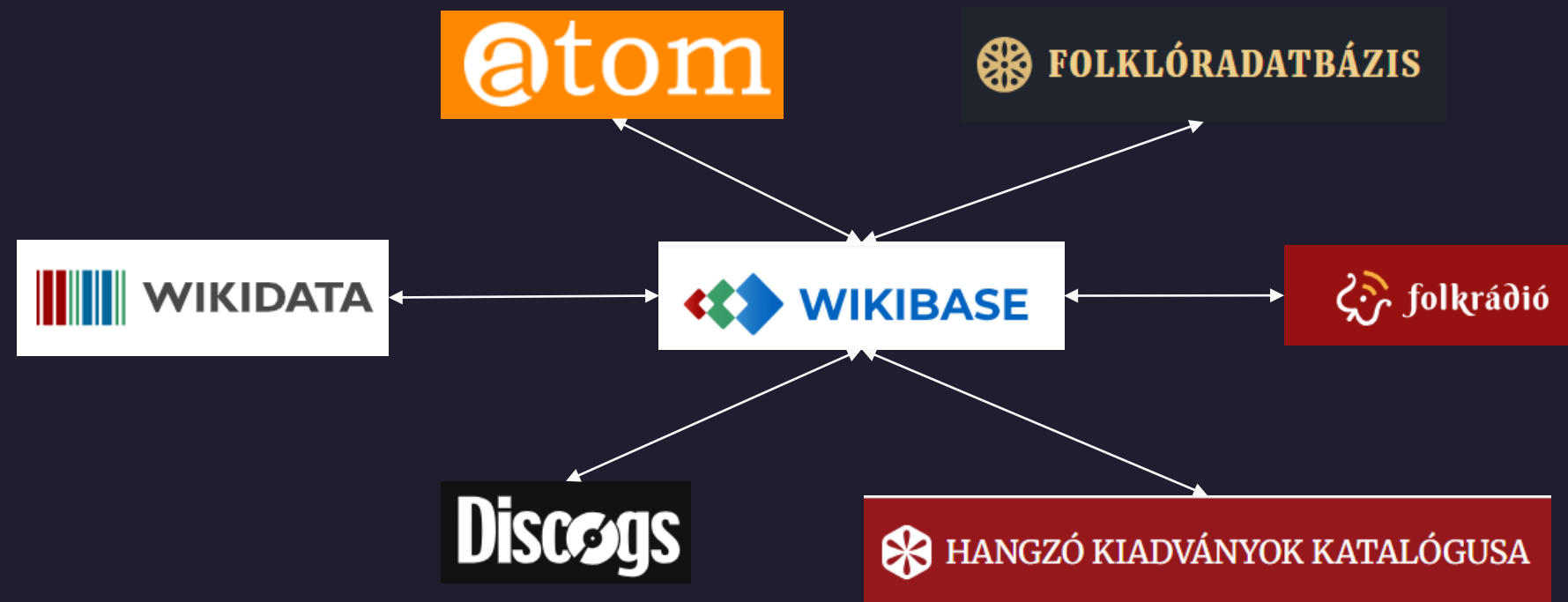
- Vízjelezés
- Adatvédelem és személyes adatok (PII) védelme
- **Jogosultságkezelés: kiterjesztett jogkezelés az alap AtoM rendszeren túl**
- Gyűjteménykezelés: múzeumok és levéltárak számára fizikai gyűjtemények nyilvántartásához
- Adományozási szerződések kezelése: adományozási megállapodások és feltételeik nyomon követése
- Kulturális örökségi eszköznyilvántartás
- Biztonsági jogosultsági rendszer korlátozott hozzáférésű iratok kezelésére
- **Kutatói portál**
- **Auditnapló**
- Több-bérlős (Multi-tenant) támogatás: több független archívum futtatása egy telepítésből
- **Records in Context (RiC): Az ICA új leíró szabványának támogatása**
- **Adatmigrációs eszközök**
- **Mentés és visszaállítás**
- Bootstrap 5 sablon: modern, reszponzív felület
- **API-hozzáférés**

Névterek és szemantikus kapcsolatok



Cél: a Hagyományok Háza szakmai adatbázisainak összekapcsolása egymással és a globális tudáshálózattal

Wikibase alapú névterek: személyek, testületek (kiadók, zenekarok, stb.), földrajzi helyek, kiadványok



Jó példa: A Wales-i Nemzeti Könyvtár SNARC projektje

The screenshot shows the main page of the SNARC project. The header includes the LGC NLW logo and navigation links. The main content area features the title 'Semantic Name Authority Repository Cymru' and a description: 'Linked database of authority records and related data for people and places in Welsh archives'. It also displays '163,297 Items.' and a link to 'Hacathon Hanes 2026'. A sidebar on the left contains various links and tools.

Magyar Zene Háza – Reprex: Hundb

The screenshot shows a Wikibase item page on the Hundb website. The item is 'Bonchidától-Bonchidáig' (Q1203), an album by Kalamajka released in 1988. The page includes a table with language labels and descriptions, a 'Statements' section with instance, title, and creator information, and a sidebar with navigation links.

Language	Label	Description	Also known as
British English	No label defined	No description defined	
magyar	Bonchidától-Bonchidáig	zenei album a(z) Kalamajka előadásában, megjelenési éve: 1988	Bonchidától-bonchidáig
English	Bonchidától-Bonchidáig	album by Kalamajka released in 1988	Bonchidától-bonchidáig

Köszönöm a figyelmet!

Zagyva Natália
zagyva.natalia@hagyomanyokhaza.hu



**HAGYOMÁNYOK
HÁZA**