

Simon András – Szabó Attila

Szemantikai háló építése az Országos Színháztörténeti Múzeum és Intézet adatbázisában

Az Országos Színháztörténeti Múzeum (OSZMI) rövid története

A 2018-as év egyik legjelentősebb múzeumi informatikai fejlesztéseként készült el az Országos Színháztörténeti Múzeum és Intézet portálja a „Magyar Színházi emlékezet megőrzése” elnevezésű projekt keretében. A munka fő célja a múzeum adatbázisában található és széles érdeklődésre számot tartó színháztörténeti anyag bemutatása az interneten. A portál készítése során sokoldalú keresési és böngészési lehetőséget kívántak biztosítani a kutatók és a laikusok számára is. A projekt megvalósításánál kulcsszerep jutott az integrált múzeumi gyűjteménykezelő rendszernek. Erre a célra a múzeummal szervezeti kapcsolatba kerülő Petőfi Irodalmi Múzeumban már több mint tíz éve használatba vett Huntéka rendszer lett kiválasztva. A Huntékába be kellett tölteni, és egy integrált adatbázissá össze kellett építeni az eddig elkülönülten fejlesztett múzeumi adatbázisokat. A betöltés, adattisztítás és integráció során az adatbázisokban található rekordokból szemantikus háló jött létre. A cikk bemutatja a munka lépéseit és a létrejövő adatbázis rendszertervének változásait. A munka során egy úgynevezett „hibrid rendszer” jött létre, mely részben a múzeumi számítógépes leltárkönyv és a könyvtári katalógus tulajdonságait ötvözi. Az adatszerkezet fel van készítve külső forrásból történő adatgazdagításra és adatszolgáltatásra idegen adatbázisok, vagy akár a Nemzeti Névtér felé is.

Tárgyszavak: színháztörténet; szakmúzeum; múzeumi gyűjtemény; tartalomszolgáltatás; tudásreprezentáció

Az Országos Színháztörténeti Múzeumot 1952 decemberében alapították azzal a céllal, hogy a kortárs színművészetet a magyar színjátszás múltjával megismertesse. Az OSZMI egyedülként kezdte el szisztematikusan gyűjteni, és gyűjti ma is a magyar színháztörténet tárgyi emlékeit. A gyűjtemény létrehozásában fontos szerepe volt a *Gobbi Hilda* színésznő által ugyanabban az évben megalapított Bajor Gizi Színészmúzeumnak, mely 1953 óta az intézmény kiállítóhelyeként működik. Az első igazgató, *Hont Ferenc* mellett *Cenner Mihály*, *Staud Géza*, *Dömötör Tekla* fektették le a hivatásos magyar színjátszás történetét dokumentáló gyűjteményezési gyakorlat alapjait. Képzőművészeti és tárgyi emlékek mellett korán elkezdődött a színházi tárgyú sajtócikkek módszeres gyűjtése a Dokumentációs tárból és számos elméleti és történeti szakkönyv, könyvsorozat jelent meg a Múzeum gondozásában. 1957-től Színház- és Filmtudományi Intézetként, majd 1959-től Színház-tudományi Intézetként működött tovább. 1970–1973

között alakult ki a ma is fennálló szakgyűjteményi rendszer, létrejött a Fotótár, a Képzőművészeti gyűjtemény, az Emléktár, a Kézirattár, a Szcenikai gyűjtemény, a Színlaptár, a Topográfiai gyűjtemény, a Hangtár. A Báltár és az 1988-ban csatlakozó Táncarchívum különgyűjteményként működnek és gyűjtik az ő színházi részterületükre vonatkozó sokféle szakanyagot. 1987-ben a Videotár létrejöttével vált teljessé a gyűjteményi struktúra. Az intézményt 1991-ben ismét átkeresztelték, ekkortól a múzeumi tevékenységet is hangsúlyozó Országos Színháztörténeti Múzeum és Intézet (OSZMI) néven szerepel, mely alapító okirata szerint országos gyűjtőkörű szakmúzeum.

A színháztörténeti muzeológia módszertanilag annyiban mindenképpen különleges terület, hogy a feldolgozás középpontjában nem annyira az egyes műtárgyak leírása áll, hanem a színházi előadásé, ami egy sajátosan strukturált eseményként ragadható meg. Az egyes gyűjteményeket összefogó

entitás tehát a színházi előadás, melyhez számos különféle dokumentum csatlakozhat. Az előadások számítógépes adatbázisának elkészítését már a nyolcvanas évek közepén elkezdték. 1985-ben a müncheni Deutsches Theatermuseummal való együttműködést követően elkészült a Bemutatótár adatstruktúrájának a gerince. 1998-ban vezették be az Ariadne rendszert, mely már nemcsak előadások adatait kezelte, hanem egy integrált múzeumi adatbázis irányába mutatott. 2008-ban az adatbázis, immár Monari néven, internetes alapra került¹. Az egyes adatbázisok – mint a bemutatótár, személyi adatbázis, audiovizuális archívum, fényképadatbázis, cikkarchívum – közti adatkapcsolatok létrehozásával összetett keresési lehetőségek nyíltak meg a kutatók, felhasználók számára. Ez itt még elsősorban a Bemutatótári rekordok bekötését tette lehetővé az egyes műtárgyakhoz, írott és képi forrásokhoz. Az adatbázis struktúrája az újabb és újabb gyűjtemények digitalizációjának és számítógépes metaadatolásának elindításával szervesen, és az egyes gyűjtemények specifikus igényeit követve alakult. A 2005–2009 közt futó Jedlik Ányos program, a 2007–2011 között végzett TACE (Theatre Architecture in Central Europe) színházépítészeti kutatás, majd a 2011–2013 között lezajlott ECLAP (European Collected Library of Artistic Performance) projekt lehetőséget teremtett bizonyos gyűjtemények és gyűjteményrészek módszeres digitalizációjára. Legutóbb pedig a Manda-program tette lehetővé, hogy a napi feldolgozáson felül jelentős mennyiségű digitalizált tartalom jöjjön létre. A legradikálisabb átállás a Dokumentációs tárban történt, már 2006-ban, amióta csak digitálisan gyűjtik a színházi cikkeket.

Az OSZMI 2017-ben a Petőfi Irodalmi Múzeum (PIM) fiókintézménye lett. Az összeolvadási folyamat részeként célszerűnek tűnt megvizsgálni a két intézmény által használt múzeumi adatbázisok harmonizációjának lehetőségét is. Bár a két múzeum gyűjtőköre sok ponton hasonló, a színháztörténet dokumentációja sajátos gyűjteményi struktúrát követel. A modern színházelmélet alapvetően mindig is arra törekedett, hogy a színházművészetet önálló, az irodalmi alapanyagtól, a drámától elkülönülő művészeti ágként határozza meg, különös hangsúlyt fektetve az előadó-művészet vizuális, zenei, testi és politikai dimenziójára. A közös működés megkezdésekor ezért mindkét intézmény vezetői számára egyértelművé vált, hogy az OSZMI múzeuminformatikai rendszerének megújítása praktikusán csak egy önálló logikát követő saját rendszerben képzelhető el. A tervezés és kivitelezés során ugyanakkor a két intézmény igen

hasznosnak tartotta egymás jó gyakorlatainak megosztását, az OSZMI rendszerének kiépítése jelentős mértékben támaszkodott a PIM évtizedes digitalizációs tapasztalataira és a Huntéka rendszer fejlesztésében elért eredményeire. Így az OSZMI múzeumi adatbázisának fejlesztésével a Huntékát fejlesztő céget, a Monguz Kft-t bízta meg.

Az áttöltési munka (melyet a Monguz Kft. munkatársai, mindenekelőtt Lengyel Mónika végzett el) során az OSZMI Monari rendszerben rögzített különálló adatbázisai tehát összeolvasztásra kerültek a a Huntéka Múzeumi Rendszerbe. Minden múzeumi leltári tétel, mely a Monariból érkezett, önálló leíró metaadatrekorddal jött létre a Huntékában, és ugyanígy minden, az eredeti Monariban tárolt szótárelem is önálló besorolási rekord lett a Huntéka adatbázisban. A különálló adatbázisokból érkező azonos szótárelemekből, a lehetőségekhez képest automatikusan végzett egységesítés során mindig csak egy besorolási rekord képződött le a Huntékában. Az áttöltési munkát e cikk egy későbbi fejezete ismerteti.

Az áttöltési munkánál a múzeum vezetése fontosnak tartotta, hogy a PIM és filiáléi a műtárgyak és dokumentumok leírásában azonos elveket kövessenek, és hogy a teljes intézményrendszer digitális adatkincse rendezett módon teljesen egyedi azonosítóval ellátva kerüljön a repozitóriumba. A mesterpéldányok biztonságos és központosított tárhelyen történő tárolásához az informatikai hálózat és a szerveres eszközpark jelentős fejlesztésére is szükség volt. Középtávú cél, hogy a PIM összes fiókintézményének különálló adatbázisokban feldolgozott tartalmi egy közös keresőben legyenek elérhetők, illetve, hogy a névtereket is közösen használhassák, építhessék. A közös névhasználat mindkét intézmény számára előnyös lehetőségeket rejt, hiszen a PIM igen gazdag, alapvetően szerzőközpontú és általános közéleti személyiségekre épülő személynévtére a színházi terület specifikus alkotóinak részletes adataival bővíthet, az OSZMI pedig „mentesül” a színházban is gyakran játszott irodalmi szerzők részletes biográfiai feldolgozásától.

Az OSZMI-ban az NKA Színházi Kollégiumának támogatásával épp az összeolvadáskor indult egy új, nagyszabású kutatási projekt is, *A magyar színházi emlékezet megőrzése* címmel, mely összetettebb adat- és médiakezelési lehetőségeket kívánt meg. A kutatás első éves időszakának egyik fontos célkitűzése volt egy olyan új színháztörténeti portál létrehozása, amely az elkészített történeti

tanulmányok illusztrációjaként, egységes felületen mutatja meg a kutatásban szereplő színházak intézményes működéséhez kapcsolódó összes forrásanyagot és a bemutatott előadások fennmaradt dokumentumait is (www.szhaztortenet.hu). A leíró adatok közti gazdag kapcsolati háló kiépítésével a portálnak összetett keresési és szűrési lehetőségekkel rendelkező felhasználóbarát felületet kell biztosítania a felhasználóknak: kutatóknak, a színházi oktatás résztvevőinek és a színházrajongóknak. A portál a színháztörténeti tanulmányok közzétételének új formáival is kísérletezik: a szövegekben hivatkozott források és médiatartalmak, a megemlített személyek és előadások adatai a szövegben szereplő linkeknek köszönhetően azonnal behívhatók.

Bár az új színháztörténeti portál csak három színház részletes történeti feldolgozását tartalmazza, beépült mögé az OSZMI által évtizedekig használt Színházi Adattár teljes adattartalma. A Színházi Adattár a magyar nyelvű színházi bemutatók átfogó adatbázisává vált, mely referenciaértékű adatforrása a színháztörténeti kutatásoknak és a színházzal kapcsolatos szakmai publikációknak. A Színházi Adattár új alapokra helyezett változata a www.szhaztortenet.hu portál részletes keresőjében (OPAC) él tovább. A változások nem ráncfelvarrás jellegűek. Ahhoz, hogy a Színházi Adattár a magyar színháztörténeti tudás átfogó szakértői rendszerévé alakuljon, nem elég, hogy a múzeum minél nagyobb mennyiségű digitális vagy digitalizált dokumentumot tegyen közzé, hanem egy gazdag metaadatrendszer kiépítésére is szükség van, a nyilvánított adatok tisztítására és minél mélyebb szintű összekapcsolására. Ha mindez megtörténik, az összetett keresések pontos, releváns és kimerítő találatokat adnak majd például egy adott drámai mű bemutatóinak számáról, időbeli és térbeli megoszlásáról vagy lehetővé válik az egyes alkotók pályaképeinek és művészi kapcsolati hálójának szemléletes felrajzolása. Ha a rendszer a színházi kritikák és szakcikk szövegében is lehetővé teszi a keresést, gépi eszközökkel előállított szógyakorisági táblák lekérésével az egyes színházesztétikai törekvések vagy fontos szakmai fogalmak kialakulásának és terjedésének feltérképezése is megvalósulhat. Egy jelentős méretű szövegtár, kiegészülve a jól metaadatolt képi és audiovizuális forrásokkal, számos kutatói „sejtés” vagy hipotézis empirikus ellenőrzésére adhat lehetőséget.

Jövőbeni tervek/jogi dilemmák

A színháztörténeti kutatás sokat vitatott, de mégis legszemléletesebb forrásait az előadás-felvételek

képezik. Egyedül a felvételek teszik lehetővé, hogy a kutató az adott előadás minden művészi aspektusáról képet kaphasson. Az OSZMI igen gazdag audiovizuális gyűjteménye elavult vagy rohamosan elavuló hordozókon van eltárolva, döntően VHS kazettákon és DVD lemezekon. A régi hordozók mellett a nagy kutatói érdeklődés is indokolja, hogy ezt a gyűjteményt a múzeum haladéktalanul digitalizálja. A *Színházi emlékezet...* kutatás keretében majdnem kétezer felvétel digitalizációja megtörtént és folyamatban van a felvételek adatbázisba való betöltése, metaadatolása. Az előadás-felvételek igen bonyolult és zömmel tisztázatlan szerzői jogi helyzete miatt a múzeum csak a saját helyszínén jogosult a felvételeket a kutatóknak megmutatni, miközben egyre határozottabb igény mutatkozik az egyetemek részéről és időnként az egyes színházak részéről is arra, hogy a múzeum lehetővé tegye a hozzáférést ezekhez a forrásokhoz. A szerzői jogi megkötések miatt az adatbázisban jelenleg csak rövid részletek tekinthetők meg a felvételekből. A színházszakmai közvélemény egyetért az előadás-felvételek elkészítésének fontosságával, de abban nincs összhang, hogy kinek a feladata lenne a felvételek elkészítését finanszírozása. A legkevésbé tisztázott kérdés, hogy ki és milyen formában biztosítaná az ezekhez való jövőbeli hozzáférést. A digitalizáció és a digitális nyilvántartás itt is számtalan új lehetőséget rejt. A digitalizáció nemcsak megkönnyíti a mindenkori új formátumokra való konverziót, hanem a kutatói igények rugalmasabb kiszolgálására is eszközöket biztosít. Az integrált múzeumi rendszer lehetővé teszi, hogy egy felvételnek egy adott részletét vagy részleteit egy adott időszakra tegyék távolról elérhetővé a kutató számára streamelt formában, míg eddig a részletek kiadásához manuális szerkesztőmunkára és többhetes átfutási időre volt szükség. Távlabbi vízió egy szakértői hálózat kiépítése, mely a NAVA pontok mintájára, regisztrált színházak, egyetemi és kutatóintézetek, esetleg könyvtárak és levéltárak saját helyszínein lehetővé teszi az audiovizuális források megtekintését.

Huntéka előzmények – személyi állományok névtérbe rendezési kísérletei két hasonló gyűjtőkörű intézményben, a Petőfi Irodalmi Múzeumban és a Hadtörténeti Múzeumban

Az olvasót talán meglepi, hogy miért tekintjük hasonlóknak a látszólag eltérő gyűjtőkörű és tudományterületet kutató három múzeumot: Hadtörténeti Intézet és Múzeum, Petőfi Irodalmi Múzeum és Országos Színháztörténeti Múzeum és Intézet. Azonban az illető gyűjteményekben nagyon nagy

számú, személyekhez köthető múzeumi tárgy lelhető fel, melyek egy-egy személy életútját kísérik végig. Ezek különféle kéziratos és nyomtatott anyagok, fényképek és személyes tárgyak egyaránt lehetnek. Egy átlagos területi múzeumban (ezek a városi és megyei hatókörű városi múzeumok), ezzel szemben igen nagy a néprajzi, régészeti, természettudományos, illetve a történeti és iparművészeti anyag aránya, melyek esetében nem ennyire központi kérdés a múzeum gyűjtőköréhez tartozó tárgyakkal kapcsolatban álló személyek adatainak és életútjának, annak szinte minden eseményének rögzítése.

Adatbázisépítés a Huntéka gyűjteménykezelő rendszerrel

A gyűjteménykezelő rendszerek eredeti küldetése a számítógépes katalógusok építése, illetve a gépi leltárkönyvek vezetése. Ezek a munkák az egyes intézményeknél évek, évtizedek óta zajlanak, a kiépült hatalmas, nagy költséggel létrejött adatvagyon további hasznosítása pedig egyre erőteljesebben felmerülő társadalmi követelmény. A gépi katalógus- és leltárkönyvépítés, különösen az egyre szabványosabb szakkatalógusok létrejöttével, önmagától szakértői rendszereket hoz létre, ahol rengeteg hozzáadott tudás halmozódik fel a tartalmi feltárás, illetőleg a különféle tárgyleírások készítése terén. A besorolási állományokból, azok folyamatos bővülésével, gazdagodásával névterek, az eleinte csak egyre részletesebb metaadatokat tartalmazó adatbázisokból pedig, teljes tartalmakat őrző és szolgáltató adatbankok fejlődnek ki. Miért is lehetnek a gyűjteménykezelő rendszerek gépi katalógusai alapjai a szakértői rendszereknek?

Mindenekelőtt hosszú évek gyűjtő és feltáró munkája során, hatalmas adattömeg gyűlik össze. A könyvtári katalógusok és múzeumi leltárkönyvek ugyanis a gyűjteményben tárolt objektumok pontos leírására törekednek az azonosítás és a visszakereshetőség céljából. Az adatbázisban rögzített tudás újrahaznosítására a fenntartó és a társadalom részéről jogos az újra meg újra felmerülő igény. Új, változó korunkban a szerepüket kereső könyvtárak hajlandók és motiváltak is, hogy ezt a feladatot felvállalják, a múzeumok pedig követik példájukat.

Az ilyen adatbázisok felépítése hatalmas szakmai feladat, az ehhez szükséges tudást amúgy sem lenne könnyű összegyűjteni, a gyűjtemények kezeleinél pedig ugyanakkor ez készen áll, hiszen a

szakma legjobb ismerői építik a gépi katalógusokat és az azokat főként a múzeumokban háttérből kiegészítő teljes tartalmakat szolgáltató adatbázisokat. Múzeumi adatbázis nem is képzelhető el teljes tartalmak, mindenekelőtt tárgyfotók megléte nélkül, mert a tárgyfotókat a nyilvántartás számára el kellett a múltban is készíteni. Ahol számítógépes leltárkönyvet vezetnek, ott a múzeum munkatársainak a fényképeket is digitalizálniuk kell, illetve újabb tételeknél már eleve digitálisan kell készíteniük.

A könyvtári és múzeumi munka a gépi katalógusok építése és szolgáltatása terén, egyre több ponton találkozunk, és egyre nagyobb a jelentősége a két intézménytípus határterületén lévő gyűjteményeknek és szervezeteknek. Mindezek következményeképpen, egy alkalmazás, mely mindkét területen használható, minden más vetélytársát megelőzően alkalmas arra, hogy a gyűjteménykezelő rendszerekből kinövő, szemantikus háló technológiára épülő szakértői rendszerek (más szóval tudásszervező rendszerek) keretétül szolgáljon. Ezzel magyarázható a Huntéka integrált gyűjteménykezelő rendszerben megvalósult, részben e cikkben is felsorolt megoldások elterjedtsége.

Előző feladat: Az összes magyar író adatbázisba rendezése

A gyűjteményi adatbázis szakértői rendszerre történő fejlesztésének úttörő, és talán máig legnagyobb volumenű példája a Petőfi Irodalmi Múzeumban épített névtér. Ez már kezdettől fogva több volt, mint könyvtári katalógus, és több mint múzeumi gépi leltárkönyvet tartalmazó adattár. A múzeum által az 1980-as évek óta épített, és mérhetetlen szellemi vagyont képviselő Access alapú adatbázisok 2006-ban, a Huntéka integrált könyvtári rendszerben való összetöltése korunkra a Magyar Nemzeti Névtér alapja lett².

A rendszerben épített adatbázisok közül az alábbi felsorolásban szereplő első három volt a legjelentősebb, legtöbb adatot tartalmazó adatbázis:

Magyar életrajzi index,
Emigráns írók lexikona,
Díjak és díjazottak,
Magyar főnemesség genealógiája,
Magyar családtörténeti adattár,
Magyar írók sírjai,
Francia becsületrend magyar kitüntetettjei,
Diplomások adattára,
Kortárs magyar írók adattára.

Az összeépített adatbázisban feltárt adatelemek köre igen széles. A könyvtári katalógusrekordok, múzeumi tárgyak, besorolási állományok, (földrajzi helyek, testületek, épületek, foglalkozások) mellett a legkiemeltebb figyelem, személyek adatait tartalmazó besorolási állomány övezte és övezi.

A könyvtári név besorolási rekord megszokott tartalmán túlmutatóan, itt kulcsszerepet kap az azonosítás, melynek keretében a pontos név, a névváltozatok, előnevek, születés és halálozás lehető legpontosabban megállapított és lejegyzett ideje és helye, valamint a foglalkozás játszik elsőrangú szerepet. A katalógizálási és leltározási feladatokon túlmenően, az egyes nevezetes személyek munkássága, művei, életútja, az általuk folytatott tanulmányok, az elnyert kitüntetések, díjak, és a mindezen információk forrásául szolgáló adatforrások is rögzítésre kerülnek. Jól látható, hogy leszámítva a létrehozott művek jegyzékét, egyik sem olyan adat, mely kifejezetten az irodalmi élet személyiségeire jellemző adat lenne, tehát a Petőfi Irodalmi Múzeum számára kialakított magyar névtér adatmodell, más területeken is érvényesíthető és terjedőben is van, különösen a megyei hatókörű múzeumok és könyvtárak körében keltve figyelmet és érdeklődést. Az így kialakított adatszerkezet, és adatrögzítési, illetve szolgáltatási elvek jelentősége tehát túlmutat e múzeum, sőt az irodalomtudomány területein is. Hasonlóan az irodalmi élet szempontjából jelentős személyeket tartalmazó névtér építése és karbantartása során felmerült problémák is általános szakmai érvénnyel bírnak.

A személynévtér építésekor rendre felmerülő szakmai kérdések

Alapvető probléma minden időben és térben tagoltan létrejött adatbázisnál a több helyről bevitt személyek azonosítása, és az azonosság megállapítása esetén a rekordok összevonása adatvesztés nélkül. Jelen cikk egyik szerzőjének, Simon Andrásnak szilárd meggyőződése szerint mindenfajta összevonás alapelve, hogy sokkal nagyobb veszély, és sokkal nehezebben javítható hiba két, valójában nem azonos rekord összevonása, mint azonos adatelemek külön rekordban való meghagyása. Ezért az azonosság kétségtelen megállapításánál és két rekord összevonásánál különös gonddal kell eljárni, mert az eredeti állapot helyreállítása (az összevont rekordok adatainak szétválasztása) szinte megoldhatatlan.

Az adatbázisokban a rekord létrehozása, annak meghatározása, hogy ha az nem köthető egy gyűj-

temény valamely fizikai eleméhez, nem egyszerű feladat. A besorolási állományokban ez általában nem egyszerű, de éppen a természetes személyeknél kézenfekvő feladat. A besorolási név megállapításánál a Petőfi Irodalmi Múzeumban követett elv, az irodalmi és tudományos, sőt a politikai életben is célszerű és jól alkalmazható. Más kérdés, hogy az éppen az e cikkben tárgyalt színház-történeti adatbázis ez alól az elv alól kivétel. A Petőfi Irodalmi Múzeumban követett gyakorlat szerint, melyet számos hazai múzeum átvett és követ, a névváltozások (asszonynevek, hivatalos névváltoztatás, írói álnevek és művésznevek, továbbá névváltozatok és eltérő névhasználat) esetében amennyiben lehetséges, akkor a születéskori anyakönyvezett név az irányadó. Asszonynevek és hivatalos névváltoztatás, illetve bizonyos esetekben művésznevek, vagy írói álnevek esetében ettől el lehet tekinteni, ha az illető megváltoztatott neve, illetve asszonyneve ismertebb, a személy irodalmi munkásságát azon fejtette ki. Jó példa erre a névváltoztatás kapcsán felvett és széles körben ismertté vált név tekintetében *Gárdonyi Géza*, illetve az asszonynévre *Illyés Gyuláné (Kozmutza Flóra)* esete. Érdekes, hogy a pedagógiai szakirodalomban legelterjedtebb az asszonynevek használata, talán azért, mert a nők szakmai munkássága itt a többi területet időben megelőzve vált jelentőssé, elismertté. Az írói álnevek, névváltozatok csak kivételes esetben lehetnek besorolási rekordok. A nem egységes névhasználat esetében (*Balassi Bálint – Balassa Bálint*), mely elsősorban a régi korok személyeire érvényes, megegyezés szerint kell valamelyik nevet választani. Az adott példa esetében a Balassi név az irányadó, de a másik névformát őrzi például az esztergomi városi múzeum elnevezése (MNM Balassa Bálint Múzeuma).

Az azonosítást nehezíti a könyvtári katalógizálás, illetve az egyes múzeumi dokumentumok, például aprónyomtatványok feldolgozása során az adatbázisba került adatszegény nevek sokasága. Sok esetben a múzeum kezelésébe került dokumentumon (nyomtatvány vagy kézirat) szerepel ugyan egy személynév, de ahhoz az illető tárgyon csak kevés információ áll rendelkezésre. A személy nevét fel kell venni, pontos azonosítására azonban a feldolgozás során sokszor nincs idő, vagy mivel kevés információ áll rendelkezésre az azonosítás bizonytalan. Ezek jelenlétét, megfelelő jelöléssel való ellátás után, meg kell tűrni az adatbázisban. Gyakran a maradéktalan biztonságú azonosításról nemcsak idő, hanem a megfelelő források hiányában is, végleg le kell mondani.

Különös, időigényes, de érdekes szakmai feladatot jelent a különböző adatforrásokból vagy különböző adatbázisokból érkező ellentmondásos adatok kezelése. Ezek legkézenfekvőbbben, gépelési hiba miatt jöhetnek létre. Javításuk után az adatmódosítás tényét a rendszerben rögzíteni és őrizni kell. Amennyiben a különbség az adatforrásokból magukból érkezik, de tévedés vagy sajtóhiba okozza, akkor a feldolgozó szakembernek kell megállapítania melyik a helyes adat, de adateltérésként a téves információt is megjelölve, és tévesnek feltüntetve, az adatbázisban a külső felhasználók számára is láthatóvá kell tenni, mert, ha a téves, vagy túlhaladott adat publikálásra került, mások is újra meg újra rá fognak bukkanni. A makacsul tovább élő és felbukkanó sajtóhibákra a legjobb, bár nem kifejezetten személynevekre vonatkozó példa az 1912-ben ott megtartott, Marokkó ügyét rendezni kívánó konferenciáról ismert Algeciras város neve, melynek hibásan gépelt változata, Algericas újra meg újra felbukkant a magyar ismeretterjesztő irodalomban, sőt az iskolai tankönyveken is. Azért is fontos a földrajzi nevek pontosságáról beszélnünk, mert kézikönyvekben ezek igen gyakran vannak hibásan leírva, és ez az egyes személyek azonosítását is megnehezíti, mert az egyes adatbázisokból jövő személyeknél különböző születési és halálhelyet fog eredményezni, jóllehet, a település azonos, csak más módon jegyezték le az adatbázisba.

Hasonló közkeletű elterjedt tévedés, hogy a Budapesten, a Gellért-hegyen található Citadella tervezője *Kasselik Ferenc* volt. Történészek (mindenekelőtt *Csapó Csaba*) közlése szerint ez téves, Kasselik Ferenc pályázata nem nyert, a tervező valójában *Ignaz Weiss* volt. Érdekes módon ez a tévedés igen elterjedt, neve a Citadellával kapcsolatban több kézikönyvben, sőt még egy az épület falán elhelyezett emléktáblán is olvasható.³

Természetesen az adatbázisban a változások nyomán követése állandó feladat, így biztosítható az adatbázis naprakészsége. Módot kell adni a közönségnek is, hogy módosítási, helyesbítési javaslatait, elsősorban saját szűkebb hazájára, községére, családjára vonatkozó adatok tekintetében megtehesse. Az egyes személyneveket, személyes adatokat tartalmazó adatbázisok nyilvánosságra kerülésével az illető személyek esetleg még élő hozzátartozói, leszármazottai felfigyelhetnek a tévesen megadott adatokra, különösen a születési és halálhelyét, illetve a születési és halálhelyét, emellett esetenként a foglalkozás, iskolák, munkahelyek sírhely tekintetében is. Az

ilyen észrevételeket érdemes fogadni, és megfelelő ellenőrzés után az adatbázisban érvényre juttatni. Helytörténeti gyűjtemények esetében az önkéntes adatközlőktől érkező információk egyenesen nélkülözhetetlenek.

Különösen ügyelni kell, a személyiségi jogok érvényesülésére. Egy névtéradatbázis a legkülönfélébb esetekben közölhet az olvasókkal érzékeny adatokat, párttagságot, munkahelyet, iskolai végzettséget, de gond lehet akár a születési év nyilvánosságra hozásával is. Ami a jelenleg érvényre jutó, és közismert rövidítéssel GDPR-nek nevezett szabályozást illeti, ez alapvetően megváltoztatja a személyes adatok kezelésének és főként nyilvánosságra hozatalának gyakorlatát. A korábbinál is nagyobb hangsúlyt kell helyezni az adatforrás azonosítására, az adatok használati jogának rögzítésére, és az adatbázis-feltörés elleni védelmére.

Mindezek a szempontok értékes tapasztalatokkal szolgáltak, egy a színészeket és színtársulatokat, színházi embereket tároló adatbázis esetében, ezért is tűnik és tűnt célszerű megoldásnak a PIM és az OSZMI adatbázisát közös elvek és logika szerint felépíteni és működtetni.

A Hadtörténeti Múzeumban szerzett tapasztalatok

A kívülről meglepheti, hogy a Hadtörténeti Intézet és Múzeum adatbázisa és gyűjteményeinek egy része mennyi hasonlóságot mutat a PIM adatbázisával, különös tekintettel a személyek nyilvántartására. A munkának az szabott irányt alapvetően, hogy az intézmény múzeumi és könyvtári gépesítési célokra is a Huntéka rendszert használta és használja, továbbá az első világháborús évforduló miatt, a szakmai és nyilvános érdeklődés közepontjába került történelmi korról ránk maradt dokumentumtömeg digitalizálása, és nyilvánosságra hozatala itt is adatbázis, és repozitóriumpépítést tett indokolttá. Az egyik legfontosabb feladat, a veszteséglisták digitális feldolgozása, illetve általában a múzeumi emléktárhoz köthető személyi adatbázis kiépítése lett.

A szokványos könyvtári munkán túlmutató nehézséget jelent a még feltárássra váró óriási adattömeg. A fennmaradt dokumentumok jó része német nyelvű, ráadásul németesen leírt, vagy átírt neveket és sokszor ma már feledésbe merült, vagy pontatlanul rögzített német településneveket tartalmaz. A munka egyelőre nem is tud itt túlterjedni az előkerült emléktárhoz digitalizálásán, a tisztí-

karhoz tartozó személyek azonosításához sokéves kutatómunkára lesz szükség.

Adatbázisba rendezhető-e az összes magyar színész, színtársulat és előadás?

Összevetve az írókkal és a katonákkal, a színészek adatainak névtérbe rendezésénél kétségtelen előnyök állapíthatók meg. Mindenekelőtt kevesebben vannak. Persze a színházi szakemberek, és a segéd személyzet, a színház működését kiszolgálók köre nagyobb létszámot jelent, mint csak maguk a színészek, de a szám még így is lényegesen kisebb, mint a PIM, vagy a Hadtörténeli Múzeum esetében. Így például a névtérbe rendezhető nevek az OSZMI adatbázisában a PIM-es szám ötödét teszik csak ki.

A színházi szakmában megszokott módon, a művészek által használt nevek legalább egy időszakban egyediek, tehát egy új, a pályára lépő színész nem használja egy még aktív kollégája nevét.

Elsősorban a színészekre, de rendezőkre, dramaturgokra is igaz, hogy egy-egy személy ismertebb és több róla az adat, mely a múzeumi emberek, kutatók rendelkezésére áll, és általában több dokumentum is kapcsolódik hozzá. Lényegesen kisebb a régebben (XX. század előtt) élt személyek aránya is.

Nehezíti a munkát ugyanakkor a testületek, szervezetek kezelése. Ezek egy hadsereg esetében annak természetéből fakadóan nagyon meghatározott struktúrát jelentenek, a színházak és társulatok esetében hasonló pontos logikát már nehéz találni. Mindamellett a színházakat és társulatokat az adatbázisban a rendszer szállítójának, a Monguz Kft.-nek munkatársai nem is akarták, elsősorban technikai, felhasználói kényelmi okokból a többi testületi név közé, közös besorolási listába betölteni. Egyrészt azért, mert a társulatok eredetileg is önálló adatállományt képeztek, és lényegesen nagyobb figyelem övezte ennek az adatállománynak a karbantartását, bevitelét, mint a többi testületnek, másrészt az adatbeviteli felületet is a társulatok esetében a testületi nevekétől eltérő módon kellett megtervezni.

A szakértői rendszerek megjelenése és szerepe a „Big Data” korában

A szakértői rendszereket (expert system) más szóval tudásszervező rendszereknek is nevezzük. A tudásszervező rendszer (Knowledge Organizati-

on System, ismert rövidítéssel KOS) a szakirodalomban igen tág fogalmat jelent. E cikkben a fogalmat a szakirodalom által többdimenziósnek nevezett információs rendszerekre szűkítjük le. Ilyennek tekinthetők az ontológiák, szemantikus hálózatok, teauruszok, és például az osztályozási rendszerek.⁴

Az információk elektronikus eszközökkel való tárolása és szolgáltatása óriási mennyiségi növekedést eredményezett, mind abszolút számokban, mind pedig az egy ember által egységnyi idő alatt elérhető adattömeg tekintetében. A mennyiségi növekedés nem feltétlenül jár együtt minőségi fejlődéssel is, főként mert a publikációk nagy része már kevesebb szerkesztési lépésen megy át, illetve az egyes publikációk létrehozására fordítható egységnyi idő arányosan egyre kevesebb lesz. Emellett a diszciplínák határainak elmosódásával egyre többször foglalkozik valaki egy olyan határterülettel, melynek nem igazán szakértője, és ilyenkor a publikációba könnyen csúsznak apró tévedések, még neves ismeretterjesztő szaklapok esetében is. Az ilyen előnytelennek tűnő változás kiküszöböléséhez nagy segítséget nyújtanak a gépi adatbázisok és adattárak, mivel ezekben a keresés hallatlanul gyors, és az adatok gépi ellenőrzése is megoldható.

E cikk egyik szerzőjének Simon Andrásnak meggyőződése, hogy fontos kihasználnunk a különféle adatbázis-kezelők, különösen a relációs adatbázis-kezelők nyújtotta előnyöket, hogy a nagy mennyiségű adatot olyan minőségben szolgáltatathassuk, mint ahogyan azt korábban, a szervezettebb publikálási, kiadási tevékenység során megszokhattuk. Az elektronikus adattárolás és publikálás előnye, hogy a javítás azonnal érvényesíthető, és a sajtóhibákat ki lehet iktatni. Nem maradnak örökre bent, mint a kinyomtatott dokumentumok esetében.

A tudástömeg megfelelő használata, a segédkönyvek további fejlődése várható, és itt egyre inkább kulcsszerepe lesz a szakértői rendszereknek, amennyiben azok valóban jó minőségű szerkesztett és karbantartott adatokat tartalmaznak.

Szemantikus háló: több mint gépi katalógus

Egy, az előző fejezetben kifejtett elvek szerint összeállított adatbázis rekordjaiból szemantikus háló építhető fel. A szemantikus háló egy tudásreprezentációs technológia, definíció szerint „irányított gráf, amelynek csúcsai (objektumok) és élei (kapcsolatok) címkézettek”⁴. Fontos megjegyezni, hogy

a szemantikus hálón két különböző dolgot értenek a szakma képviselői. A fenti definíció a kifejezés azon értelmére igaz, melyet más szóval, „szemantikai hálónak” is nevezhetünk. E sorok szerzője ebben az értelemben használja a szemantikus háló kifejezést. Ugyanakkor a „szemantikus háló” kifejezés alatt sokan szemantikus világhálót, szemantikus webet értenek⁵.

Az adatokat, a háló csomópontjait és kapcsolataikat, a háló éleit természetesen nem elegendő csak begépelni. A felvitt tartalmak minőségénél, azok használhatóságánál kulcsszerepe van a szabványosításnak, és az adattartalom szegmentációjának. A háló csomópontjait képező adatrekordoknál döntő a csomópontok között létrehozott linkek megléte is, így irányított és definiált kapcsolatok, azonosított, címkézett és validált (igazolt tartalommal rendelkező) csomópontok hálózata jöhet létre a metaadatok alkotta adatbázisból. Mindennek működtetésére, tárolására, szolgáltatására és őrzésére egy gépi katalógus a legalkalmasabb, ahol a nagy adattömeg miatt az azonosíthatóságnak és a visszakereshetőségnek is nagy jelentőséget kell tulajdonítanunk.

Jóllehet korábbi publikációinkban már foglalkoztunk a kérdéssel, nem haszontalan kitérnünk a szakértői rendszer minősítésénél kulcsfontosságú szempontok meghatározására, hiszen ezek nélkül aligha beszélhetünk valódi szemantikus háló építéséről.

Első szempont a szabványosság:

Az egyes adattartalmakat, azok szerkezetét és jellegét – szerencsés esetben – szakmai konszenzuson és hatósági megerősítésen alapuló szabványok határozzák meg. Mindezek mind a könyvtári, mind a múzeumi, mind a levéltári szakmában jelen vannak, és a gépesítéssel, az adatbázisok távoli hozzáféréseivel, és a gépi katalógusok integrálásával egyre részletesebb szabályozást tesznek lehetővé, egyre általánosabb érvényűek lesznek, és egyre szélesebb körben alkalmazzák is őket.

Második szempont a szegmentáció (tagoltság):

Minél tagoltabb, minél inkább elkülöníthető, önállóan is értelmezhető alapegységekre van bontva egy adat, mind műszaki, mind pedig szakmai tekintetben, annál inkább használható önállóan, átvihető másik adatbázisokba, biztosítható elkülönült szempontként való jelenléte összetett keresések során, és karbantartása, javítása, illetve áttöltése más adatbázisokba is kevésbé munkaigényes, tehát olcsóbb.

Harmadik szempont a címkézettség:

Az egyes adattartalmak akkor tekinthetők önálló egységnek, ha saját címkével (tag-gel) rendelkeznek. A címkézettségénél fontos az egyértelmű azonosítás, a tévedést, félreértelmezést kizáró kapcsolat az adatelem és címkéje között, és az hogy az egyes címkék tartalmilag, és szakmailag jól meghatározottak legyenek. Igazából az adatbázis minősége és használhatósága szempontjából másodlagos, hogy maga a címkézettség formailag milyen elemekből áll, tehát hogy az illető adatbázis esetleg a MARC-szabvány, vagy a LIDO⁶, címkéit alkalmazza-e. Minél tagoltabb az adatszerkezet, és minél sokoldalúbb a címkézettsége, annál inkább képes adatokat veszteségmentesen tárolni és őrizni, illetve annál kevesebb munkával készült, és annál jobban hasznosítható outputokat tud kiadni akár továbbhasznosítás, akár adatcsere részére.

Negyedik szempont a validáltság (igazoltság):

Korunk az egyre kevésbé ellenőrzöttek terjedő információ kora. Ezért kulcsszerepe van a megfelelő tekintélyű intézményeknek, szervezeteknek, hogy az általuk a nyilvánosság felé is közölt tartalmaikat igazolják. Az igazoltságnak az egyes elemi egységekre is ki kell terjedniük.

Ötödik szempont: a kapcsolatok irányítottsága és definiáltsága (meghatározottsága):

Nemcsak az adattartalmakat, hanem kapcsolatait is címkézni kell, és ennek éppen úgy meg kell felelnie a szakmai és minőségi szempontoknak, mint az adattartalmak címkéinek. A kapcsolat nem mindig szimmetrikus, ezért az irányítottságot is jelölni kell.

Szótárelem, értéklista, besorolási rekord – egy funkció fejlődése

A szemantikus háló létrehozásánál már mindenképpen a meglévő adatbázisokra támaszkodik minden intézmény. Az adatbázisok fejlesztésénél fejlődést figyelhetünk meg. Ezt a fejlődési folyamatot legkönnyebben a háttérállományok kezelési módja tekintetében kísérhetjük végig. Az első adatbázisok még szótárelemeket használnak, illetve bizonyos esetekben még ma is használják ezeket. (Amikor kevés releváns adatelem van, ezek ritkán változnak, és ugyancsak ritkán jönnek létre újak.) A szótárelemekből a fejlődés egy bizonyos fokán értéklisták jöttek létre. Ezek már nem egyszerű kitöltési segédletek voltak, hanem egy háttérállománnyal dolgoztak, mely esetben a módosított adattartalmak automatikusan az egész adatbázisra

zisban módosultak. A következő lépcsőfok a besorolási rekord, mely esetben az értéklisaelem már önálló rekord az adatbázisban, mely legalább a hasonló besorolási rekordokkal és meghatározott kapcsolatok szerint a leíró rekordokkal is összekapcsolható. Igazi szemantikus háló akkor jön létre, ha a besorolási rekordokat más típusú besorolási rekordokkal, illetve teljes tartalmakat feltáró rekordokkal is irányítottan és definiáltan össze lehet kapcsolni.

Egy gondolkísérlet: szemantikus háló további építésére használható alkalmazás fejlesztésének lehetősége

Egy szakirodalmi adatbázisban az adatelemek közötti kapcsolatok felderítése és rögzítése az adatok nagy mennyisége és az emberi munka alkalmazásának korlátozott lehetősége miatt vonzó lehetőség. Az emberi értelem kiváltása még illúzió, egyelőre csak a lexikális adatállomány könnyebb elérhetőségéről és kezeléséről, vagyis javításáról és naprakészen tartásáról lehetne szó.

Az emberi tudás, az egyes adatok közötti kapcsolatok adatbázisba rendezésével felmerül a kérdés, elképzelhető-e, hogy egy számítógépes alkalmazás önállóan hozzon létre megadott logikát követve további kapcsolatokat egy adatbázisban? Elképzelhető-e olyan alkalmazás, mely az új kapcsolatokat (melyeket emberi asszociációknak is megfeleltethetünk) felkínálja jóváhagyásra egy szerkesztőnek, és a felkínált kapcsolatok elfogadottsága, vagy el nem fogadottsága alapján képes tanulni, és egyre jobb, egyre inkább elfogadható kapcsolatokat felkínálni? Elképzelhető-e olyan alkalmazás, mely emberi kontroll nélkül is már alacsony hibaszázalékkal működik, sőt az adatbázis kapcsolati hálójának ismeretében új asszociációk létrehozására képes anélkül, hogy a szabályokat a program tervezője meghatározná? Egy ilyen tanulásra képes rendszert akár már mesterséges intelligenciának is nevezhetünk.

Mindettől persze, a jelenlegi szakértői rendszereink még nagyon messze vannak. Az adatok között további kapcsolatok építése természetesen algoritmizálható. Példákat adhatunk az adatbázis automatikus javításának lehetőségeire, de ezek helyes működéséhez az alapul szolgáló adatbázisnak igen sokrétű és jó minőségű adatokat kell tudni szolgáltatni. Továbbá az ellenőrzést igen sok esetben le kell folytatni, hogy az alkalmazás választásainak helyességét tesztelhesük.

Alább egy kiragadott példával jelenítjük meg a szemantikai háló automatikus bővítésének lehetőségét az OSZMI adatbázisában. Ahhoz persze, hogy egy ilyen program valóban működni tudjon, és automatizmus segítségével hozzáadott tudást építsünk be egy adatbázisba, nagyon sok, nagyon jó minőségű adatot kell szolgáltatnunk.

Szakértői rendszerben feltárhatjuk Schiller Don Carlos c. drámája, illetve Verdi hasonló című operája között a kapcsolatot, illetve, hogy Schiller drámáját *Orániai Vilmos* egy (a valóság alapot egyébként teljesen nélkülöző) Habsburg ellenes politikai röpirata ihlette. De ha mindezeket egy adatbázisban feltárjuk, még csak a lehetőségét teremtettük meg annak, hogy egy program a rendszerben további kapcsolatokat hozzon létre. Amennyiben a szakértői rendszert kibővítjük *Pierre-Augustin* Caron de Beaumarchais „Figaro házassága” c. művével, és *Mozart* hasonló című operájával, és még néhány hasonló példával, például a francia színműíró „A Sevillai borbély” c. darabjával, és *Rossini* hasonló című operájával, mindegyik esetben megállapítjuk a kapcsolatot a színdarab és az opera librettója között, és ezután a rendszer egy hasonló műpárost (hasonló című színdarab és opera, ahol a színdarab keletkezési dátuma megelőzi az operáét, és a szereplők legalább részben hasonló nevéek), és a kapcsolatot jóváhagyásra felkínálja egy szerkesztő irodalmárnak. Majd annak elfogadása után a sikeres kapcsolat létrehozásának tényét a program regisztrálja az esetleges sikertelen kapcsolatokkal együtt, összefüggést talál a sikeres és a sikertelenül létrehozott kapcsolatok megállapítási módja között, és egyre ritkábban hoz létre később az emberi ellenőr által el nem fogadható kapcsolatokat, majd elegendő adat birtokában olyan szabályszerűségek alapján talál kapcsolódó rekordokat, mely szabályt eredetileg az algoritmusba nem építettük bele. Így ez már a mesterséges intelligencia határán levő megoldás lenne. Persze ez az intelligencia is igazából „természetes”, hiszen az adatbázis karbantartásával megbízott informatikusnak kell egy algoritmust az illető szakemberekkel együttműködve megalkotnia (ez a nehezebb) és programoznia, (ez a könnyebb).

Simon András a Petőfi Irodalmi Múzeum munkatársaival közösen az adatbázisba került névrekordok közül próbálta meg kiválasztani azokat, melyek valójában ugyanarra a személyre vonatkoztak⁸. Ekkor több hullámban kerültek felderítésre a hasonló rekordok, de nemcsak a rekordok azonoságára vonatkozó döntés, hanem az azonosnak

vélt rekordok kiválasztása (igazából a két rekord közötti kapcsolat megállapítása) is mindig emberi döntés eredménye volt. Ez a kiválasztás újabb és újabb szempontok figyelembevételével, akár történhetett volna gépi erővel is. A munka tanulságai felhasználhatók lehetnek a jövőben egy sokkal nehezebb és időigényesebb feladat végrehajtásához, az adatbázisrekordok mű besorolási rekordokhoz való rendeléséhez.

Jelenleg a közgyűjtemények adatbázisai, különösen a könyvtári adatbázisok a bibliográfiai leírás elvei alapján kiadványokat tárnak fel, a kiadványokból jönnek létre a könyvtári katalógusokban a bibliográfiai rekordok, amelyekhez a művet mint besorolási rekordot rendelhetjük. Ez a mai gyakorlat szerint a tartalmi feltárást végző könyvtáros feladata, tehát emberi erővel történik, és a másolós katalógizálás elterjedésével, az illető tudást már a könyvtárak egymástól átvehetik. Mindamellett a kiadványok a legtöbb könyvtárban a rekordok túlnyomó részénél még nincsenek mű besorolási rekordhoz rendelve. A művek és a bibliográfiai rekordok egymásnak való megfeleltetése, amennyiben a művek és azok fordításai (kifejezési formái), illetve különböző hordozón való megjelenései (pl. nyomtatott és hangoskönyv) egy adatbázisban rendelkezésre állnak. Ezekkel a kiadványokkal a szerzők, a mű címe, esetleg egyéb adatok, például tárgyszavak alapján már meg lehet feleltetni. Ez történhet a jövőben akár programmal is, mely a fentebb ismertetett logika alapján képes lehet tanulásra, tehát a sikeres és sikertelen döntések körülményeinek figyelembevételével minél több sikeres döntést meghozni.

A művek egymás közötti kapcsolatainak a megjelenése ennek a munkának egy további folytatása lehetne. Különösen hasznos ez a munka az OSZMI adatbázisa esetében, ahol a művek (opérák, színdarabok és egyéb művek) sokszor nagyon erősen átdolgozásra kerülnek. Itt természetesen a szakember döntése nem megkerülhető, de a döntés előkészítésére gépi eszközöket el lehet képzelni.

Az OSZMI színháztörténeti projektje

A projekt kezdetei, a rendszerterv változásai

A magyar színházi emlékezet megőrzése kutatói projekt 2017 márciusában indult, célja az volt, hogy új szempontok szerint gondolja újra az OSZMI által irányított színháztörténeti kutatásokat,

számot vetve a XX. század második felétől nagy mennyiségben fennmaradt audiovizuális dokumentációval. A papíralapú publikációhoz képest az internetes közlés sokkal gazdagabb képi és audiovizuális forrásanyag közlését teszi lehetővé. A múzeumba ugyanakkor a fennmaradt forrásoknak csak egy töredéke kerül be és félő, hogy egy színházi intézmény megszűnésekor ezek a dokumentumok örökre eltűnnek. A projekt elsődleges feladata tehát az volt, hogy egységes módszertani keretet dolgozzon ki egy adott színház teljes körű történeti dokumentációjára, megpróbálva legalább digitális formában begyűjteni a számos különböző helyszínen (színházaknál, önkormányzatok raktárában, levéltárakban, városi múzeumokban, magánszemélyeknél) őrzött színházi forrásokat. A kutatás szempontrendszerét és a projekt első körében részt vevő színházainak listáját egy kilencfős szakértői bizottság dolgozta ki. Így került sor a budapesti (már megszűnt) Bárka Színház, a nyíregyházi Móróc Zsigmond Színház és a zalaegerszegi Hevesi Sándor Színház történeti feldolgozására. *Lakos Anna* projektvezető kutatási terve a feldolgozás prioritásaként több szempontot meghatározott: szerepeljen vidéki és budapesti színház is, lezárt történetű és élő, változásban lévő színházi műhely is a kutatásban. Az első körben feltárt színházak közti sok történeti párhuzam mellett a szakmai bizottságnak praktikus szempontokat is figyelembe kellett vennie: a száz-kétszáz évet felölelő történeti színházi műhelyek elmélyült feltérképezésére a projektnek nem lett volna elegendő kapacitása az első etap szűkös időkeretében, amikor a módszertani keretrendszer és az informatikai háttér kidolgozására is sort kellett keríteni. Az egyes helyszíneken színháztörténészekből, dramaturgokból, művészeti munkatársakból álló csoportok végezték a kutatómunkát, akik számos oral history interjút is rögzítettek. A projekt internetes portáljának elsődleges feladata a kutatás során megszületett történeti tanulmányok és szerkesztett interjúk szövegének publikációja volt, kiegészítve a begyűjtött fényképek, videofelvételek, levéltári dokumentumok és különböző promóciós anyagok (plakátok, műsorfüzetek) digitalizált változataival. Az OSZMI projektvezetői számára a tervezés korai szakaszától kezdve világos volt, hogy a színháztörténeti tanulmányok és a különböző képes illusztrációk közti kapcsolatot a színházi előadások adatbázisának kell megteremtteni, de a konkrét informatikai-adatszerkezeti megoldásokat az OSZMI által megbízott informatikai partnerrel, a Monguz Kft.-vel közösen kellett kidolgozni.

A rendszerterv változásai

A projekt alakulását, „fejlődését” jól szemlélteti a megvalósulás két éve alatt a rendszerterv hogyan változott meg. Kezdetből egyértelmű volt, hogy ez a sajátos adatbázis egy az eddigi múzeumi és könyvtári gyakorlatban nem túl nagy hangsúlyt kapó adatféleségen az „előadáson” (alapvetően színházi előadáson) alapult. Középpontba tehát egy más adatbázisokban segédállományként kezelt rekord került⁹.

Az előadás mint önálló rekord kezelése nem idegen a Huntékától, sőt a Huntéka könyvtári rendszertől sem. Elsősorban megyei könyvtárak kapják azt a feladatot, hogy az illető megyeszékhelyen bemutatott előadásokat azok valamennyi adatával (szereplők, közreműködők, előadás helye, ideje, körülményei) feldolgozzák, őrzik és közölik, mint fontos helytörténeti adatot. Egy könyvtári rendszerben egy város néhány eseményének feltárásakor azonban az előadás leírásánál jelentkező problémák, mindenekelőtt az előadásnak mint olyannak a meghatározása nem merül fel meghatározó kérdésként. Az előadás tökéletes meghatá-

rozását természetesen nem lehet megadni, de egyezményesen az OSZMI gyakorlatát követve a bemutatókhoz kötődik. Az előadás első megvalósulása, premierje kitüntetetten leírt entitás az OSZMI adatbázisában. Egy-egy előadás természetesen többféle szereposztással is megvalósulhat, előadható több helyszínen, és idővel meg is változhat, esetleg már annyira, hogy önálló előadásként kezelhetjük. (Ugyanez igaz lehet a bibliográfiai leírás során a művekre. Egy-egy átdolgozott újabb kiadás esetenként már annyira megváltozik, hogy nem is lehet ugyanannak a műnek tekinteni.) A múzeum gyakorlata szerint ezért a bemutatók mellett a felújításokat is már, mint önálló előadásrekordokat vették fel. Az előadások természetesen művekhez köthetők, de ezek mögé az eredeti adatbázisban még nem volt mű besorolási állomány rendelve.

Az alábbi ábrán jól látszik a színdarab és az opera egymástól elkülönült feltártsága, (az opera szerzőiként a szöveggönyv írói lettek megjelölve).¹⁰ Az elkülönülést az ábra második és harmadik oszlopának tartalma szemlélteti (1. ábra).

Don Carlos	[Dráma három felvonásban]	Schiller, Friedrich	bemutató	1990.02.09
Don Carlos	[Drámai költemény két részben]	Schiller, Friedrich	bemutató	1992.12.04
Don Carlos	[Tragédia két részben]	Schiller, Friedrich	bemutató	1994.09.30
Don Carlos	[Dráma három felvonásban]	Schiller, Friedrich	bemutató	1987.02.27
Don Carlos	[Dráma három felvonásban]	Schiller, Friedrich	bemutató	1984.03.16
Don Carlos	[Tragédia két részben]	Schiller, Friedrich	bemutató	1974.10.04
Don Carlos	[Dráma]	Schiller, Friedrich	bemutató	2001.03.09
Don Carlos	[Színmű három felvonásban]	Schiller, Friedrich	bemutató	1969.02.22
Don Carlos	[Két részben]	Schiller, Friedrich	bemutató	2005.11.18
Don Carlos	[Dráma három felvonásban]	Schiller, Friedrich	bemutató	1959.03.13
Don Carlos	[Dráma három felvonásban]	Schiller, Friedrich	bemutató	1963.01.25
Don Carlos	[Dráma három felvonásban, tizennyolc képben]	Schiller, Friedrich	bemutató	1956.04.27
Don Carlos		Schiller, Friedrich	bemutató	2008.03.21
Don Carlos		Schiller, Friedrich	bemutató	2011.11.26
Don Carlos		Schiller, Friedrich ;	bemutató	2012.11.18
Don Carlos	[Tragédia]	Schiller, Friedrich	bemutató	2013.01.12
Don Carlos	[Drámai költemény öt felvonásban]	Schiller, Friedrich	felújítás	1857.11.30
Don Carlos	[Drámai költemény öt felvonásban]	Schiller, Friedrich	bemutató	1881.02.28
Don Carlos	[Drámai költemény öt felvonásban]	Schiller, Friedrich	bemutató	1916.11.10
Don Carlos	[Drámai költemény öt felvonásban]	Schiller, Friedrich	bemutató	1932.02.26
Don Carlos	[Drámai költemény öt felvonásban]	Schiller, Friedrich	bemutató	1942.11.14
Don Carlos	[Opera három felvonásban]	Méry, Joseph ; Lacle, Camille du	bemutató	1988.03.04
Don Carlos	[Opera három felvonásban]	Méry, Joseph ; Lacle, Camille du	bemutató	1988.06.27
Don Carlos	[Opera két részben]	Méry, Joseph ; Lacle, Camille du	bemutató	1988.08.13
Don Carlos	[Opera három részben]	Méry, Joseph ; Lacle, Camille du	bemutató	1994.02.19
Don Carlos	[Opera három felvonásban]	Méry, Joseph ; Lacle, Camille du	felújítás	1983.04.08
Don Carlos	[Opera három felvonásban]	Méry, Joseph ; Lacle, Camille du	felújítás	1981.09.24
Don Carlos	[Opera négy felvonásban, két részben]	Méry, Joseph ; Lacle, Camille du	bemutató	1998.06.11
Don Carlos	[Opera két részben]	Méry, Joseph ; Lacle, Camille du	bemutató	1998.12.11
Don Carlos	[Opera három felvonásban]	Méry, Joseph ; Lacle, Camille du	felújítás	1979.11.10
Don Carlos	[Opera négy felvonásban, hét képben]	Méry, Joseph ; Lacle, Camille du	bemutató	1988.02.24
Don Carlos	[Opera három felvonásban]	Méry, Joseph ; Lacle, Camille du	bemutató	2000.03.10
Don Carlos	[Opera öt felvonásban]	Méry, Joseph ; Lacle, Camille du	bemutató	1979.07.03
Don Carlos	[Opera három felvonásban]	Méry, Joseph ; Lacle, Camille du	új betanulás	1974.01.23
Don Carlos	[Opera négy felvonásban]	Méry, Joseph ; Lacle, Camille du	új betanulás	1973.11.30
Don Carlos	[Opera három felvonásban]	Méry, Joseph ; Lacle, Camille du	felújítás	2001.12.15
Don Carlos	[Opera három felvonásban, hét képben]	Méry, Joseph ; Lacle, Camille du	felújítás	1969.03.21
Don Carlos	[Opera három felvonásban]	Méry, Joseph ; Lacle, Camille du	bemutató	2003.10.17
Don Carlos	[Opera három felvonásban]	Méry, Joseph ; Lacle, Camille du	bemutató	1980.07.04
Don Carlos	[Opera négy felvonásban]	Méry, Joseph ; Lacle, Camille du	bemutató	1967.08.13
Don Carlos	[Opera három felvonásban]	Méry, Joseph ; Lacle, Camille du	bemutató	1963.11.22
Don Carlos	[Opera három felvonásban]	Méry, Joseph ; Lacle, Camille du	bemutató	1958.05.23
Don Carlos	[Opera négy felvonásban]	Méry, Joseph ; Lacle, Camille du	bemutató	1961.06.04
Don Carlos	[Opera három felvonásban]	Méry, Joseph ; Lacle, Camille du	új betanulás	2009.12.18
Don Carlos	[Opera öt felvonásban]	Méry, François Joseph ; Du Locke, Camille	magyarországi bemutató	1868.03.14
Don Carlos	[Félig szcenírozott előadás két részben]	Méry, Joseph ; Lacle, Camille du	bemutató	2007.02.03

1. ábra A Don Carlos mint opera és mint színdarab az OSZMI adatbázisában

Az OSZMI eredeti rendszere több adatbázisból állt, melyek a múzeum gyűjtőköre és tevékenysége szerint különültek el. Az adatbázisok részletes ismertetésére a cikk későbbi részében kitérünk, illetve azt az 5. ábrán bemutatjuk. Ahogyan a PIM-ben a személyek, a Hadtörténeti Múzeumban a katonai alakulatok, itt az előadások képezik a teljes adatfeltárás fő vezérlő elvét, és valamennyi esetben a rögzített információk köre bőven túlmutat a hagyományos könyvtári katalogizálás, illetve múzeumi tárgyleírás területén.

Az adatszerkezet bonyolultságát jól szemléltetik az elkészült rendszertervek. A rendszerterveket e cikk két szerzője, és Lengyel Mónika, a Monguz Kft. munkatársa készítették a projekt során. A rendszertervek feladata elsőként annak megállapítása volt, hogy alkalmas-e a Huntéka gyűjteménykezelő rendszer adatszerkezete az OSZMI adatbázisának befogadására, majd a konverzió tervezésekor, és a betöltési munka befejezésekor készült még egy-egy rendszerterv, utóbbi már a Huntékában kialakult adatstruktúra szemléltetésére.

A második az eredeti rendszerterv, tehát a rendszerterv első változata látható, mely a projekt kezdetén a szóbeli megbeszélések után alakult ki.

Az ábra közepén az előadás látható, az azt meghatározó, leíró adatokkal, adatfészekkel. Ehhez kapcsolódik a társulat, melyet akkor még intézményként (testületi névként) fogtunk fel. A társulattól megkülönböztettük a színházat mint intézményt, és attól ugyancsak megkülönböztettük a fizikai játszóhelyet. Egy színházon belül ugyanis működhet több társulat, és sok színházépület van, mely csak befogadó színház, illetve egy előadást eljátszhatnak nemcsak színházépületben, hanem művelődési házban, koncertteremben, vagy akár egy szabadtéri színpadon is. Emellett az egyazon színházépületben levő különböző termek külön játszóhelynek minősülnek.

Ettől függetlenül nyilván vannak tartva a színházépületek is. Ezeket, tehát a kifejezetten színházi célokat szolgáló épületeket egyébként egész Európára vonatkozóan, igaz, nem teljes körűen, egy prágai központban szerkesztett adatbázis tárja fel. Erről az adatbázisról cikkünk befejező részében részletesebben is megemlékezünk.

A múzeumban őrzött rengeteg féle dokumentum egy része a társulathoz kötődik. Ide tartoznak az éves jelentések, statisztikák, szerződések, igazgatói pályázatok, próbarendek, társulati fotók, interjúk, évadértékelő riportok. Ezek alapvetően művek,

de itt a múzeumi szemlélet érvényesült, így a hordozók alapján megállapított dokumentumtípusok kerülnek feltárássra, illetve írhatók be az adatbázisba. A hordozók köre igen széles, aprónyomtatványok, adattári dokumentumok, kéziratok, eredeti forrásértékű és adattári fotók is találhatóak közöttük. Természetesen amennyiben a teljes tartalom digitalizálva van, vagy mint újabban ez már természetes, már digitálisan született, akkor médiarekordként az is csatolható. A bizalmas anyagokat a nagyközönség elől természetesen el kell rejteni.

Az előadáshoz közvetlenül kapcsolódnak a Produkció (*előadásfotó, felvétel, jelmezterv, díszletterv, rendezői példány, próbanapló, szövegkönyv*) és a Disztribúció (*szórólap, reklám, színlap, műsorfüzet*) fogalommal jelölt dokumentumcsoportok, melyek a Társulathoz közvetlenül köthető dokumentumokhoz hasonlóan viselkednek. A Recepció (tanulmány, visszaemlékezés, kritika, interjú, nézői vélemény) csoportba sorolt dokumentumok az eredeti terv (2. ábra) szerint részint az előadáshoz, részint annak alacsonyabb rendű egységeihez, az Évadokhoz lettek volna kapcsolva. Az Évadoknak lett alárendelve az alkalmi bemutatók sora. Ezek valahol, valamikor máshogyan lettek előadva, általában a bemutatótól különböző szereposztással. A Személyi nyilvántartásnál a személyek a szerepeikkel kötődtek az előadáshoz, ahol a szerepet a rendszer tervezése során e cikk szerzői és a rendszer tervezői önálló besorolási rekordként képzeltek el. Az eredeti, még elméleti síkon kigondolt szerkezetben a szerep a műhöz kötődött, és az volt kapcsolva az előadáshoz. Ennek a négyes kapcsolásnak a műszaki megvalósítása az alkalmazás megjelenítő felületein nem lenne kézenfekvő, de természetesen megoldható lett volna. A személyek így kapcsolatban állhatnak egymással és a társulattal is.

A projekt alakulását jól szemlélteti a rendszerterv többszöri megváltozása. A konverziós munka megkezdésekor az adatbázist megismerve alakult ki az újabb rendszerterv. Ennek ábrája igen bonyolult lett, és a szemléltetett változatban a személyek és adataik nincsenek is jelen. Szembetűnő különbség még az, hogy az intézmény itt az előadással egyenrangú entitás. Az előadás intakt egységet képez, nem bomlik évadokra és alkalmi bemutatókra. Az egyes dokumentumtípusok fészekük szerint maguk is gráfokba vannak rajta rendezve (3. ábra)¹¹.

A negyedik ábrán a tényleges megvalósulás látható, legalábbis a mostani állapot, az OSZMI eredeti

Monari rendszerben tárolt adatbázisainak a Huntéka múzeumi és könyvtári rendszerbe történt konverziója után. Természetesen a mostani állapot sem tekinthető tökéletesen befejezettnek.

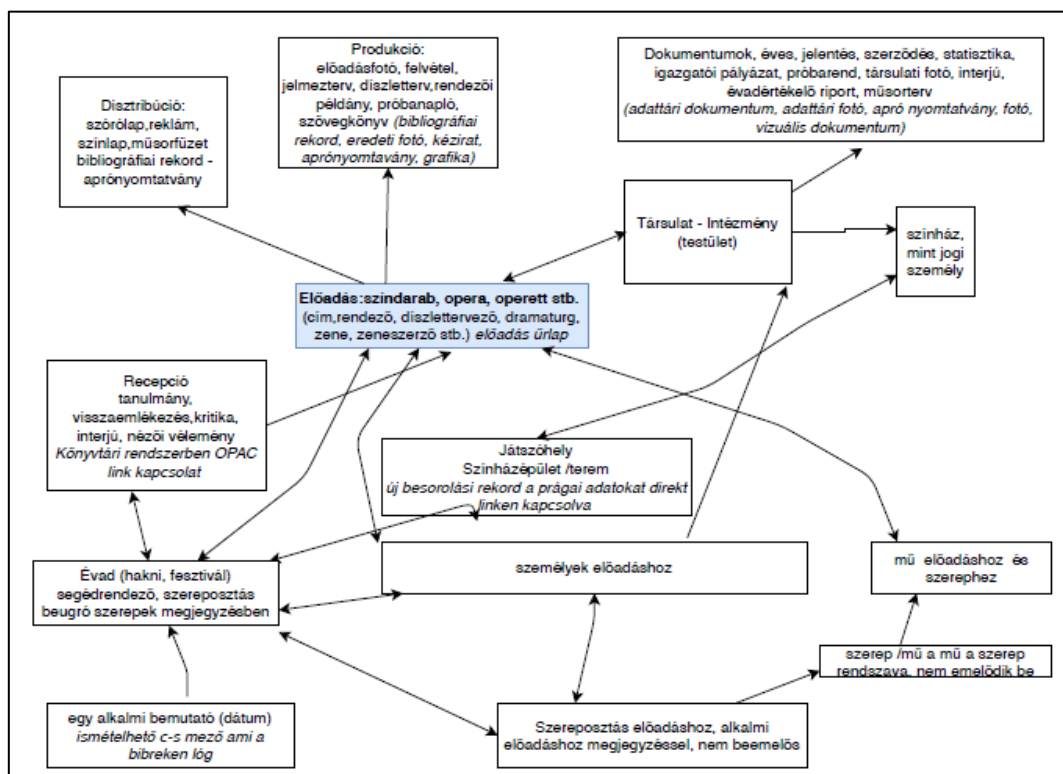
A negyedik ábrán jól látható, hogy az előadás ismételtelen egyértelműen és kizárólagosan a közép-pontban van. Ehhez kapcsolódik a mű, és részben ahhoz, részben közvetlenül az előadáshoz az előadásfotó, az audio- és videofelvétel, a jelmezterv és díszletterv, illetve a rendezői példány. A Disztribúció és Recepció adatcsoport közvetlenül az előadáshoz kötődik, illetve ismét jelen van az évad és alkalmi bemutató mint az előadáshoz kapcsolódó egység. A személyek a szereposztáson keresztül kapcsolódnak az előadáshoz, a szerep ahhoz, mint kisegítő háttérállomány kötődik. A Társulat kötődik változatlanul a személyekhez, illetve a színházhoz mint szervezethez, és a játszóhelyhez. A dokumentumok mind a társulathoz, mind a színházhoz kapcsolódhatnak.

Három közgyűjteményi típus szintézise

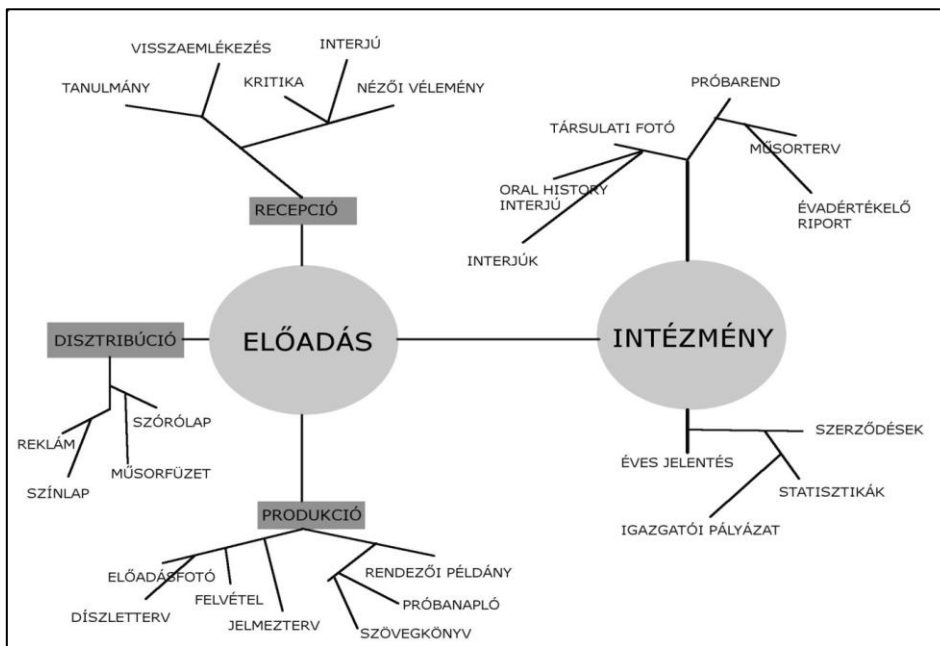
Az áttöltendő adatbázisok köre mutatja, milyen színes adatállománnyal kellett dolgozni a Huntéka

konverzió során a Monguz Kft. munkatársának (5. ábra).

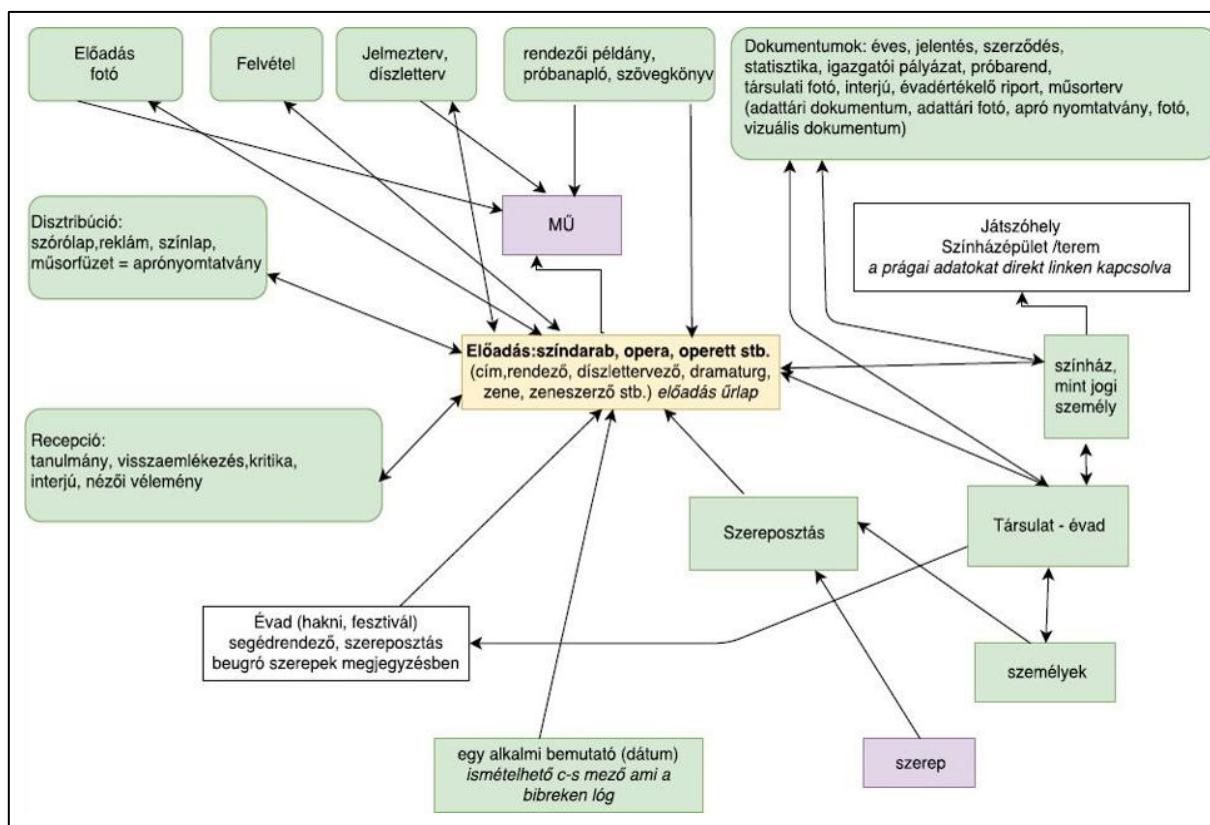
Az előadásokból az előző adatbázisban eleve hat különböző nyilvántartás létezett, amelyek egymással semmilyen kapcsolatban nem voltak, illetve elkülönítve épültek. Az Adattár tartotta nyilván a legtöbb magyar nyelvű színházi bemutatót, ezt erre szakosodott munkatársak fejlesztették szinte naprakészen a színházaktól kapott adatok, színlapok és internetes források alapján. A színházi témájú cikkeket szemlélő könyvtárosok már ezekhez, az adattárosok által felvitt előadásokhoz csatolták a feldolgozott cikkeket. Ezzel szemben bizonyos speciálisabb gyűjtemények külön előadásnyilvántartást építettek. Ilyen volt a Kézirattár és a Képzőművészeti tár, ahol számos korai, az Adattárban még fel nem dolgozott előadás regisztrálására volt szükség, vagy a Táncarchívum, ahol kissé eltérő logika szerint dolgozták fel az előadásokat. Néhány nyilvántartás az adatbázis valamely korábbi fejlődési állapotának maradványa volt. Az új rendszerben határozott szándékunk volt ezeknek a párhuzamosságoknak a megszüntetése, az esetleges redundanciák kiszűrése úgy, hogy mindeközben biztosítani lehessen az egyes műfajok és alterületek sajátos igényeit.



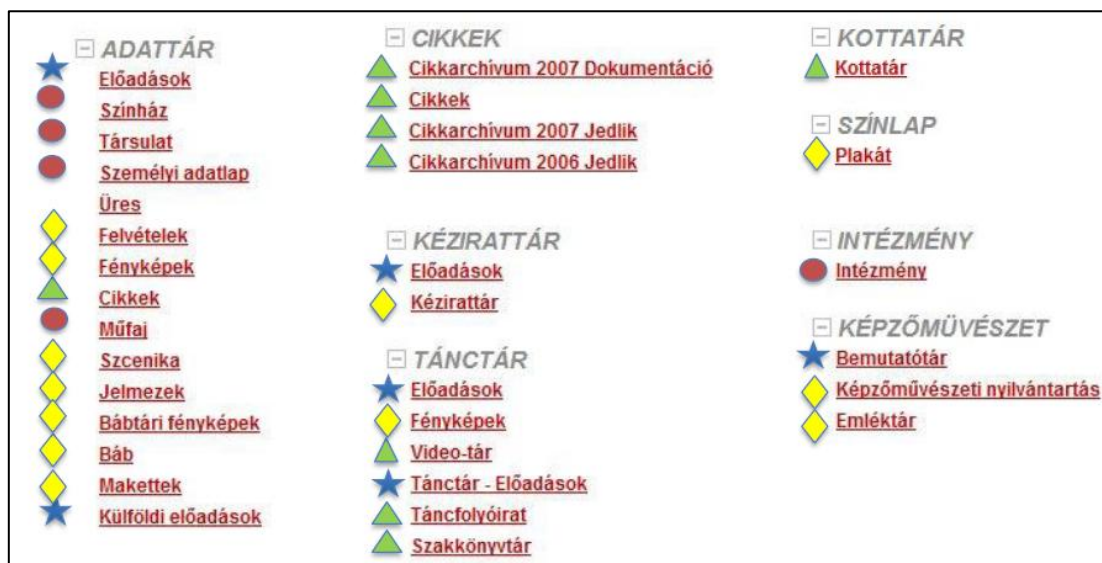
2. ábra Az OSZMI adatbázisának rendszerterve a projekt indulásakor
Simon Andrásnak az alkotása



3. ábra Az adatbázis rendszerterve a konverzió kezdeti időszakában
Szabó Attila műve



4. ábra Az adatbázis rendszerterve a konverzió elkészültekor
Lengyel Mónika a Monguz Kft. munkatársának a műve. (Közlés Lengyel Mónika engedélyével.)



5. ábra **Az OSZMI adatbázis-szerkezete**
Lengyel Mónika, a Monguz Kft. munkatársának illusztrációja

A teljesség kedvéért meg kell említeni, hogy az OSZMI-nak saját könyvtára van, mely a múzeumban elkülönült egységként már évek óta a Huntékát használja integrált könyvtári rendszerként. Mindamelllett az OSZMI múzeumi adatbázisában is voltak könyvtári rekordok, a könyvtári rendszerek megszokott bibliográfiai, példány- és besorolási rekordjaival. Ezek közül a cikkek, a feltárás és nyilvántartás tekintetében ugyan könyvtári rekordoknak tekinthetők, de tartalmilag a múzeumi tárgyakhoz kötődtek, ezért a múzeumi adatbázisba lettek betöltve. Egyéb könyvtári jellegű dokumentumok leírásai is voltak ebben az adatbázisban, de ezeket közös megegyezéssel a könyvtári katalógusba töltöttük át.¹²

Az OSZMI eredeti adatbázisa, mint múzeumi rendszer és mint számítógépes leltárkönyv, a múzeumi gyakorlat szerint a tárgyak leírását és leltárkönyvi adatait tárolta és szolgáltatta, a hozzájuk kapcsolt képekkel együtt. Ezekből épült fel az OSZMI Huntékában tárolt múzeumi tárgyleíró adatbázisa, mely a korszerű logikával elkülönítve és egymáshoz kapcsolva tárolja az adminisztratív metaadatokat és a leíró metaadatokat.

Az áttöltési munka, (melyet a Monguz Kft. munkatársai, mindenekelőtt Lengyel Mónika végzett el) során az OSZMI Monari rendszerben rögzített különálló adatbázisai tehát összeolvasztásra kerültek a Huntéka Múzeumi Rendszerbe. Minden múzeumi leltári tételt, e cikk egy későbbi fejezete ismerteti.

Az áttöltés után az alábbi múzeumi gyűjtemények jöttek létre (1. táblázat). Egy múzeumi gyűjteménybe tartozó leltári tétel a mai felfogás szerint megköveteli az őrzendő fizikai tárgy létét. A digitálisan született (born digital) múzeumi dokumentumok (például eredeti forrásértékű fotók) leltározási szabályainak kidolgozása még a múzeumi szakma jövőbeni feladata.

Az OSZMI eredeti adatbázisának a könyvtári katalóguson és a múzeumi leltárkönyvön túlmenő részét a szokványosan információs rendszerhez (névterek, tezauruszrendszerek, információs portálok) tartozó adatok tették ki, középpontban és az egész adatbázis középpontjában egy segédállománnyal, a már említett Előadás adataival.

1. táblázat

Gyűjtemények az OSZMI Huntéka múzeumi adatbázisában

Gyűjteménynév a Huntékában	Szakleltárkönyv típusa ¹³
Audio-Vizuális Archívum	Vizuális dokumentum
Bábtári fényképek	Eredeti forrásértékű fotó
Bábtári tárgy	Iparművészeti tárgy
Emléktár	Történeti dokumentum
Fotótár	Eredeti forrásértékű fotó
Jelmez gyűjtemény	Iparművészeti tárgy
Képzőművészeti	Képzőművészeti tárgy
Kézirattár	Kézirat
Kézirattár kottatára	Kotta mint múzeumi tárgy
Szcenikai gyűjtemény	Képzőművészeti tárgy
Színházi emlékezet	Kézirat
Színlaptár	Történeti dokumentum
Táncarchívum	Eredeti forrásértékű fotó
Táncarchívum	Vizuális dokumentum

Önálló bibliográfiai rekordként kerültek az adatbázisba azok a színházzakmai szempontból érdekes cikkrekordok, melyek mögött fizikai leltári tétel (sem a múzeumi leltárkönyvben, sem a könyvtári katalógusban) nem állt, és nem áll egyéb rendelkezésig a jövőben sem. Ugyanakkor, mivel a színházi kutatás igen fontos forrásait képezik a színikritikák, interjúk, az OSZMI szakkönyvtárának külön részlege foglalkozik a színházi témájú lapszemlével, melynek forrásai egyre inkább a digitális publikációk fele tolódnak el. A munkát végző kollégák nemcsak bibliográfiai leírást végeznek, hanem a cikkek teljes szövegét digitálisan lementik és médiatartalomként az adatbázisba illesztik. Ugyan számos nyomtatott és digitális szakfolyóiratnak, napilapnak van saját repozitóriuma, de ezeket sokszor átalakítják, a linkhivatkozások elhalnak vagy megváltoznak. A fejlesztés kezdeti szakaszában felmerült a Cikkarchívum rekordjainak az OSZMI által régebb óta használt könyvtári Huntékába való áttöltése az újonnan kialakított múzeumi helyett, de ezt, bármennyire logikusnak tűnik profiltisztasági szempontból, szakmai megfontolások miatt (egyelőre) el kellett vetni. A könyvtári Huntéka adatbázis ugyanis nem fér hozzá az előadások nyilvántartásához, a produkciós adatbázishoz, amely, ahogy fentebb kifejtettük, a színházi gyűjteményezés alapját képezi. Az alapvetően könyvtári gyakorlatot követő bibliográfiai leírás utolsó lépéseként ugyanis a feldolgozó feladata, hogy hozzácsoportolja az adott cikket az adott színházi előadás (vagy színházi intézmény) rekordjához. A digitális világban így „kontinuitást” nyer az 1952-

ben elindított gyűjtési gyakorlat: a Cikkarchívum évtizedeken keresztül szerzőnként és bemutatónként rendezett mappákba gyűjtötte az újságkivágatokat, illetve külön mappákba kerültek az egyes színházi alkotókra vagy általában egy színházi intézményre vonatkozó írások.

Az új leíró struktúra kialakításakor az OSZMI és a Monguz Kft. egyaránt törekedett a gyűjteményi redundanciák felszámolására, következetesen ragaszkodva ahhoz az elvhez, hogy az egyes forrástípusok azonos úrlapon, azonos gyakorlatot követve kerüljenek leírásra. Ugyanakkor a gyűjteményi struktúra sajátosságait nem lehetett figyelmen kívül hagyni: a színházi területen kívül az OSZMI két jól definiálható különgyűjteménnyel rendelkezik, a Bábgyűjteménnyel és a Táncgyűjteménnyel, amelyek saját tárgyi szakanyaggal, cikkgyűjteménnyel, audiovizuális gyűjteménnyel, kisnyomtatványtárral és kézirattárral rendelkeznek, kutatói bázisuk is erősen specializált. Bár a „próza” színház, bábszínház és táncszínház alkotói között sok (és egyre több) az átjárás, a névtér tisztításánál jól látható e három terület elkülönülése: a tánc nemzetközi beágyazottsága igen jelentős, és ez az adatok szintjén is világosan megmutatkozik. Míg a tánc esetében a koreográfussal kell számolni, mint a rendező mellett/helyett álló kiemelt szerzővel, addig a bábszínházban a bábtervező, bábkészítő szerepe igen meghatározó. Ezeket a sajátosságokat a rendszer minden szintjén érvényesíteni kell. A digitális leképezést tovább bonyolítja, hogy ezekben a különgyűjteményekben

egy leltárkönyvet sokszor több szakanyag leltározására is használtak: a bábtári leltárkönyvek például egyaránt tartalmaznak kéziratokat, eredeti fotókat, belépőjegyeket, szórólapokat. Bár a digitális feldolgozás lehetőséget ad az egy leltárkönyvben szereplő különböző szakanyagok eltérő úrlapokon történő feldolgozására, ha a cél a majdani akkreditáció és a kézi leltárkönyv-kitöltés elhagyása, akkor meg kell fontolni az adatbázisból történő leltárkönyvnyomatás szempontjait is. A könyvtári (cikkek, videók, kották, folyóiratok) dokumentumokat leíró rekordok, a múzeumi leltári tételeket (fényképek, jelmezek, makettek, bábok, adattári dokumentumok) feltáró rekordok, e két közgyűjteményi ág határán levő, tehát akár a múzeumi, akár a könyvtári gyűjteményekbe elvben felvehető tételeket (kéziratok és az intézménytörténeti dokumentumok) leíró rekordok, a betöltött teljes tartalmak, így a képek és szövegek, a besorolási adatállományba felvett szerzők, testületek illetve az egész struktúra középpontjában álló előadások így komplex egységet képeztek és képeznek. Mivel a szerepek és a művek az eredeti adatbázisban még szótárként sem voltak feltárva, idő hiányában a Monguz Kft. munkatársai nem hoztak létre a konverzió során egyelőre besorolási rekordot a szerepekből, és jórészt a művekből sem.

Az áttöltés során létrejött adatbázisban 46 féle adat-típust (rekordtípust) különböztetünk meg, és közel 1000 féle elemi egységben vannak az adatok önálló címkével ellátva tárolva. Ez utóbbiak részben megfeleltethetők az ismert MARC-formátum mezőinek és almezőinek, de nem mind, mert a Huntékában lévő sok adatféleség egy részére nincs a MARC szabványban adatmező definiálva.

Két világ határán, a múzeumi és könyvtári elemeket vegyítő „hibrid” rendszer problémái

A múzeumi és a könyvtári működés gépi kiszolgálását elméletben egy közös adatbázissal és alkalmazással is el lehet látni. Különleges esetekben a Huntéka forgalmazásakor a Monguz Kft. munkatársai ezt is vállalják. Ennek feltételei: vagy a múzeumi, vagy a könyvtári állomány a másikhoz képest legyen nagyon kicsi.

A múzeumi állomány legyen kifejezetten olyan jellegű, mely a könyvtári munkában is jól kezelhető, tehát a könyvtári adatleírási szabványt alkalmazva is legyenek a tárgyak jól leírhatók. Álljon döntően nyomtatott vagy kézíratos, papíralapú (2D

– kétdimenziós) dokumentumokból. A könyvtár ne kölcsönözzön, legalábbis ne külsősöknek. Ezek a feltételek itt nem álltak fent, ezért a két adatbázist külön tartotta meg a múzeum a konverzió során. Ezzel együtt, mivel maradtak könyvtári jellegű dokumentumok a múzeumi rendszerben, továbbra is együtt kell tudni élni, érvényesíteni a könyvtári és a múzeumi adatleírást kiszolgáló rendszerek szempontjait könyvtári és a múzeumi adatfeltárási logika egy adatbázisban történő alkalmazása esetében, szükséges az eltérő tárgykezelési szemléletek összebékítése. A múzeumi tárgyak egyediek, és legalábbis elméletben pótolhatatlanok. Ezért egy leíró metaadatrekordhoz mindig csak egy adminisztratív metaadat tartozhat. Két múzeumi leltári tárgy még akkor sem tekinthető azonosnak, ha egyforma, tehát a nyomtatott könyvhöz hasonlóan gyáripari eljárással, sorozatgyártással készült. A megszerzés körülményei, és a tárgy fizikai megjelenése, állapota ugyanis ilyen esetben is különbözhetnek. Az egyébként egyforma, de ténylegesen rekordszinten önálló leírással rendelkező tárgyak között a kapcsolatot belső kereszthivatkozással mutatjuk ki, jelezve például, hogy a kiállításon az egyik tárgy a másikkal felcserélhető.

A különböző (könyvtári és múzeumi) szabványoknak való együttes megfelelést az egymás mellett létező kétféle típusú adatleíró úrlappal kell megoldani. Bizonyos dokumentumtípusoknál, mint az aprónyomatvány vagy a kézirat, mindkét úrlaptípus az intézmény rendelkezésére áll, de persze az OSZMI-ban ezek esetében mindig csak a múzeumi típusú úrlapot használják. A múzeumi úrlapon ugyanis nem is lehet rekordot leltári szám (és ezzel együtt természetesen, adminisztratív metaadatrekord) nélkül létrehozni, és egy leíró metaadathoz mindig csak egy adminisztratív metaadat tartozhat. Ez olyan, mintha egy könyvtári adatbázisban minden bibliográfiai rekordhoz csak egy példányrekord tartozhatna. Hagyományos könyvtári rendszerek, melyek leltárkönyv szemléletűek voltak, egyébként működtek így.

Ezzel szemben a korszerű rendszerekben a könyvtári úrlapon egy bibliográfiai rekordhoz több példányrekord is tartozhat, és létrejöhetnek bibliográfiai rekordok példányrekord nélkül is. A feltárható adatok köre is rekordféleségenként más, mert a leíró felületeknek meg kell felelniük nemcsak az egyes tárgyféleségek tulajdonságainak, hanem a könyvtári szabványok mellett számítógépes múzeumi adatleírást szabályozó rendeletnek is¹⁴.

A konverzió fontosabb lépései

A konverzió során nem egyszerűen szabványos múzeumi és könyvtári adatok kerültek át egy meghatározott csereformátum létrehozása útján egyik rendszerből a másikba, hanem a több önállóan fejlődött adatbázis összetöltésével és az adatok sokrétű kapcsolásával, illetve rekordok létrehozásával egy adatháló jött létre az új rendszerben.

Megtörtént tehát az adatbázisok átforgalmazása egy MARC leírás címkéi (tagjai) által meghatározott elemekből és az elemek kapcsolataiból álló szemantikus hálóból.

Az adatbázisok összeépítése során meg kellett állapítani a többféle helyről érkező értéklistas elemek azonosságát, és azokból besorolási rekordokat kellett létrehozni. Mivel több kisebb, sok esetben részben azonos jellegű tartalommal rendelkező adatbázis volt, ezért az értéklistasokat közös besorolási állományokká kellett összefésülni.

A leíró metaadatok esetleges azonosságát szerencsére nem kellett vizsgálni, mert valamennyi önálló múzeumi tárgyat írt le. Hasonló esetekben könyvtári adatbázisok összetöltésénél a bibliográfiai rekordok listáit mindig össze kell építeni (megergálni).

Az adattisztítás lépéseit nem lehetett elhagyni. Amennyiben besorolási rekordoknál jöttek létre vélelmezhetően egyformák, mivel egy-egy rekordnál kevés információ állt rendelkezésre, ezért a feldolgozó munkatársakra maradt a feladat, hogy a hűtőkarbantartás-funkciójával, az azonosnak tekinthető besorolási rekordokat összevonják. Kapcsolataikat a Hűtőkarbantartás e funkciója már az összevonás műveletét követően összeépítette. A szótáras mezők¹⁵ esetében adattisztítást kellett végrehajtani, mely a teljesen egyforma tartalommal túl a gépelési hibák eredményeképpen létrejött felesleges szótárelemeket is felszámolta.

A rendszerben levő rekordokat, mint minden Hűtőkarbantartás rendszerben négy alaptípusra lehet osztani. Ezek: a bibliográfiai rekord (leíró metaadat), példányrekord (adminisztratív metaadat), besorolási rekord (tárgyszó), illetve médiarekord. (Zárójelben mindig a múzeumi használatban meghonosodott kifejezést adtuk meg).

Rekordtípusok a konverzió után létrejött Hűtőkarbantartás rendszerben

Média rekord

- Kép
- Borítókép

- Tárgyfotó
- Illusztráció
- Eredeti forrásértékű fotó
- Adattári fotó
- Rendezvényen készült felvétel
- Digitalizált eredeti metaadat-hordozó
- Teljes szöveg
- Tárgyleírás
- Digitalizált kiadvány
- Digitálisan született kiadvány
- Audiovizuális anyag

Bibliográfiai rekord

- Könyvtári katalógustétel – több példány is lehet
- Múzeumi leltári tétel – egy leírás egy adminisztratív egység
- Szakirodalom rögzítése, olyan bibliográfiai rekord, mely mögött nem áll könyvtári tétel, még az intézmény által digitálisan birtokolt sem.

Besorolási rekord

- Természetes személyek
- Jogi személyek
- Sorozatok
- Biológiai és geológiai rendszertani egységek (állat és növényfajok, ásványok nevei)
- Földrajzi nevek
- Technikák, stílusok, kultúrák
- Történelmi korok
- Anyagok és megnevezések
- Művek
- Társulat
- Színház mint intézmény

A besorolási rekordok esetében meg kell jegyezni, hogy mint múzeumi rendszer, a Hűtőkarbantartás ugyan mindegyik létrehozását felkínálja, gyűjtőkörök miatt az OSZMI adatbázisában nem mindegyik van valóban jelen, például a biológiai egység az OSZMI-ban egyelőre egyáltalán nem. Másrészt a társulat és a színház más múzeumi és könyvtári adatbázisokkal ellentétben fentebb már kifejtett – elsősorban történelmi okok miatt – önálló besorolási rekordként van felfogva.

Szótárelemek

Számos adat szótárelem maradt a múzeum előző múzeumi és könyvtári rendszerből való áttöltés után is:

- Nyelvek
- Országok
- Technikai – műszaki adatok
- Dokumentumtípusok
- Közreműködés jellege
- Utalónál utalás jellege

- Személyeknél iskola, nem, vallás, életút időpontjai, nemzetiség
- Díjaknál díjfajta és díj/kitüntetés/fokozattípusa,
- Felirat típusa, helye és nyelve,
- Esemény jellege,
- Technikai adatok, műfaj, műszaki jellemzők,
- Mérésnél mérés dimenziója (magasság, hossz stb.) és mértékegysége

Későbbiekben a szerep mező is szótáras lesz.

Nevek karbantartása

Személynevek rendezése, rendszerezése természetesen a PIM-hez képest kevésbé komplex feladatot jelentett. A nevek kisebb száma, a színeszek könnyebb azonosítása, a színházi élet egyéb szereplőinek, rendezőknek, koreográfusoknak, szövegkönyvszerzőknek nagyobb ismertsége, illetve az adatbázis egységesebb volta miatt az adatbázisban szereplő személyek (leszámítva a szerep mező tartalmát) könnyebben voltak azonosíthatók.

Alapvetően a PIM-ben elfogadott elvek jutottak érvényre itt is, azzal a kiegészítéssel, hogy a művésznevek itt annyira ismertek és elterjedtek, hogy ezeket tekintjük besorolási névnek. Míg egy, az írókat nyilvántartó rendszerben pl. *Rejtő Jenő* a besorolási rekord, és *P. Howard* „Lásd utalóval” van hozzákapcsolva, addig a színházi adatbázis nyilvántartásában *Grósz Dezső* (a közismert színész születési neve) helyett *Garas Dezső* szerepel besorolási rekordként a Huntéka rendszerben és a születési név van a „Lásd még” utalóban.

A személyek köre, a Színháztörténeti Múzeum adatbázisában természetesen a színészeknél és rendezőknél sokkal szélesebb, a közreműködők miatt. Egy előadásnál az ismertebb személyek, színészek rendezők, koreográfusok, dramaturgok, zeneszerzők, zenei előadók mellett mások, így díszlet- és jelmeztervezők is szerepelnek. Jelentős a külföldi vendégművészek száma is. Egyes részterületek, mint például a táncművészet, nemzetközi beágyazottsága jelentős, nagy a művészi mobilitás, ezért ezek a gyűjtemények nagy külföldi névállomány nyilvántartását követelik meg.

A személyeknél a testületek köre itt szűkebb, de itt nehézséget jelent az alapegység megállapítása. Sokszor nem egyértelmű, mikor tekinthetünk két entitást azonosnak, a szervezetek névváltoztatásai, átalakulásai miatt. A leghelyesebb, ha egy szervezet létének elismerését megpróbáljuk a

nyilvántartásba vételhez kötni. Ettől persze sok esetben el kell térni, különösen akkor, ha egy szervezet részeit is önálló besorolási rekordként akarjuk felfogni. Ez persze az egységesség rovására megy, a testületi név adatbázis az OSZMI Huntéka rendszerében igen sokrétű, részben ezért is lettek a társulatok és a színházak önálló besorolási állományba rendezve.

A testületi név mint entitás meghatározásánál tehát, jóllehet a Nemzeti Levéltár útmutatását igyekeznek a Monguz Kft. munkatársai követni, és erre bátorítva az ügyfeleket, így az OSZMI és a PIM szakembereit is, a hivatalosan bejegyzett szervezeteket tervezik testületnek tekinteni, de gyakorlati szempontból ettől számos esetben el kell majd térni. Ráadásul ez az elv már csak a jövőre nézve érvényesül, az eddig bevitt testületi nevek ilyen jellegű ellenőrzésére nincsen mód, ennek nincs is értelme. Testületeknél problémát jelenthetnek a névváltozások, amikor nem történik meg új szervezet bejegyzése, de maga a szervezet lényegesen megváltozik, illetve olyan részszervezetekkel is gond lehet, melyek nem bejegyzett testületek, de mégis önálló testületként szükséges őket kezelni. Jó példák ezekre: felsőoktatási intézmények tanszékei, cégek osztályai, hatóságok vagy egyesületek alacsonyabb szintű szervezeti egységei, melyek rögzítése egy múzeum gyűjtőkörétől függetlenül indokolt lehet.

Mindezek miatt a testületek kezelését általában nem szerencsés túlbonyolítani, és az adott adatbázis esetében a kiemelt fontossággal bíró szervezeteket önálló rekordban érdemes feltárni. Ez történt jelen esetben is, ezért került a Társulat és a színház különálló besorolási állományba.

Az OSZMI adatbázisának betöltésekor a társulatot és a színházat tehát a testületi név besorolási rekordoktól elkülönített listában, de besorolási rekordként tároljuk. Minden múzeumi adatbázis tervezésekor el kell fogadni, hogy bizonyos, az illető múzeum gyűjtőköre szempontjából fontos adatot a többi múzeumtól eltérő módon kell kezelni. Különösen igaz ez a szakmúzeumokra. A PIM adatbázisában ilyen adat az épület, melyet leíró metaadat rekordként tároljuk, a Természettudományi Múzeum Huntéka adatbázisában az embertani információk képeznek teljesen önálló besorolási állományt, míg érdekes megjegyezni, hogy a magyar gyakorlattól teljesen különbözően a gdanski (danzigi) tengerészeti múzeum adatbázisában a hajók földrajzi fogalmakként vannak felfogva (mint lelőhely/gyűjtőhely). Ezek a példák is mutatják,

hogy nem egyedi eset, hogy az OSZMI esetében bizonyos adatfészeségek kezelésekor kivételt tetünk. A leíró rekordcsere, a múzeumi „közös katalogizálás” a múzeumok esetében egyelőre úgysem része a közeli jövőképek, a múzeumi adatcsereformátumok (LIDO, és Dublin Core) tartalma pedig annyira elszegényített, leszűkített, hogy ott az adatfelfogásból eredő különbözőségek nem fognak egyelőre gondot okozni.

Gépi és kézi utómunkálatok

A szereposztás mező szétvágása

Különleges feladatot jelentett a konverzió során a szereposztás mező feldolgozása. Ebbe a mezőbe az eredeti rendszerben be lett gépelve az illető előadás szereposztása, úgy ahogyan az a színlapon látható volt. Ez az információ önmagában is értékes, a kereshetőség biztosítására az adathálót fel kellett építeni, ezért az eredeti begépelte szöveg megőrzése mellett, azt az adatbázis-kezelő eszközeivel szétdarabolva, a személyeket, és a szerepeket a megfelelő mezőkbe be kellett tölteni, a személyneveket be kellett sorolni, a bevitt neveket normalizálni kellett, az egyéb még további információt közlő bejegyzéseket, pl. (f.h. =főiskolai hallgató) külön fel kellett tüntetni. A felesleges központosítási jeleket a besorolási és szótár rekordból el kellett távolítani.

Az előadások meghatározása és egységesítése úgy történt, hogy azokat cím szerint azonosította a Monguz Kft., a konverziót végző munkatársa. Az eredeti adatbázisok létrejötténél jól kinyomozhatóan az történt, hogy az egyes adatbázisok egymásból lettek átmásolva, egyik a másikból jött létre, és e sajátos törzsfajlódást végigkövetve, meg lehetett állapítani az egyes adatbázisoknál a kezdetkor létrejött állományokat. Ezek egységesítése így már egyértelmű feladat volt. A továbbiakban az egyes adatbázisok között már semmiféle adatszinkron nem volt, ott a nevek alapján kellett az azonosságot megállapítani. A személyneveknél az összevonas alapja a név és a születési dátum volt, ugyan csak megoldás lehetett ez a testületeknél is, ugyanis itt a legtöbb esetben csak nagyon kevés információ került rögzítésre.

Érdekesképpen megmutatjuk a szerep mező egységesítése után a leggyakrabban feltüntetett szerepneveket.

Előfordulás szám | Szerep

429	/	Pincér
377	/	Anya
377	/	Rendőr
365	/	Orvos
285	/	Pap
281	/	Apa
277	/	Anna
256	/	Mária
243	/	Bíró
241	/	Péter
288	/	Éva

A következő felsorolásban láthatjuk a legtöbb előadásban szereplő színészneveket. Ezek többségénél a gyakorinak tűnő és színész művésznév-ként alkalmasan hangzó vezeték és keresztnév kapcsán az adattisztítást a Monguz Kft. részéről végző, és ezt a cikket szerzőként jegyző Simon András okkal gyanította, hogy több személyt is jelölnek ezek a nevek.

Előfordulás szám | Név

390	/	Kiss László
371	/	Horváth Ferenc
330	/	Tóth József
304	/	Kovács Gyula
276	/	Horváth József
268	/	Nagy Zoltán
263	/	Szabó István
250	/	Somló István
247	/	Nagy István
246	/	Tóth Béla
246	/	Fehér Tibor
243	/	Molnár László
241	/	Kovács István
234	/	Fillár István
233	/	Kovács János

Külső adatbázisok használata, az adatgazdagítás lehetőségei

A szemantikus háló akkor teljes, ha kimutat az adatbázisból, nem kezeli azt zárt rendszerként. Hosszú távon tervezi a múzeum a PIM adatbázisával való szinkronizációt, rekordok kölcsönös átmozgatása lehetőségének megteremtését, és

közös kereső működtetését. Emellett kapcsolat létrehozását az európai színházi épületeket feltáró Prágában szerkesztett adatbázissal.¹⁶

Ezt a felületet tanulmányozva láthatjuk, hogy ez az adatbázis sem igazi szemantikus háló, hanem egyelőre még csak egy adattár. Érdeemes lenne majd ezeket a megoldásokat továbbfejleszteni. Az egyes színházépületek, és a hozzájuk tartozó intézmények csak meg vannak említve, nincsenek több nyelven megnevezve, sem az épületek, sem az intézmények, sem a településnevek, még akkor sem, ha többnyelvű intézményről van szó.

Emellett a PIM-hez hasonlóan, a múzeum szakemberei keresik a kapcsolatot az ismert külső adatforrásokkal például személynevek esetén¹⁷.

Egy személyhez az adatbázisban is érdemes rögzíteni a külső adatforrások azonosítóját. Egyes, főként külföldi kisebb múzeumi rendszerek egyébként nem is tárolnak besorolási adatokat, hanem csak az illető adatra vonatkozó külső linket.¹⁸

Egy külső linkre jó példa az alábbi viaf-ra mutató URL, mely Garas Dezső nevéhez kapcsolódik:

https://viaf.org/viaf/69139892/#Garas,_Dezso%CC%8B

Az adatot láthatóan a PIM adatbázisából töltötték fel. A PIM-es webes keresőben ugyanez az adat így olvasható:

<https://opac-nevter.pim.hu/record/-/record/display/manifestation/PIM55455/4d90ab38-f4ea-4b1c-b6b2-1bc323cc95c7/solr/0/24/0/1/authorOrder/ASC>

Ugyanez a rekord az OSZMI webes felületén is elérhető:

<https://szinhaztortenet.hu/record/-/record/OSZMI1572/solr>

Az automatikus adatgazdagítás esetén adattartalmakat automatikusan is megfeleltethetünk külső adatforrásoknak. Mivel ez igen nagy hibalehetőséget jelent, egyelőre a kézi adatgazdagítás lehet megfelelő megoldás. Köztes megoldásként kínálkozik az esetleges Europeana feltöltés után, az ott már elvégzett adatgazdagítás eredményeinek hasznosítása. A külső adatbázisok rekordjainak adattartalmával való összevetés során létrejött kiegészített adattartalmakat visszatölthetjük a mi saját adatbázisunkba. Az Europeana karbantartóinak megvannak az eszközeik az adatgazdagításra, és mivel hosszú ideje nagy erővel folytatják

ezt, saját eljárásaik megbízhatóságát jelentősen növelték.

Közzététel az interneten

A magyar színházi emlékezet projekt számára készülő portál fontos szempontja volt, hogy igazán felhasználóbarát legyen, és ne csak a részletes keresést, a szakértő kutatást támogassa, hanem a színházrajongó, az adott társulatok előadásai és története iránt érdeklődő látogató számára is izgalmas tartalmakat kínáljon. Egy adott dokumentum megtekintésekor ajánljon sok szoroson vagy lazábban kapcsolódó tartalmat, amelyekre kattintva, egyfajta vezérfonálként, elinduljon egy izgalmas színháztörténeti barangolás. Mivel az elsődleges célközönség mégiscsak a színházi szakma, a színháztörténeti kutatók, az is fontos volt, hogy a bekerült dokumentumokat a lehető legrészletesebben, az OSZMI szokásos múzeumi gyakorlatát követve legyen feldolgozva. Hiszen épp a bőséges leíró adatok teszik lehetővé, hogy az egyes dokumentumok között gazdag és valódi kapcsolati háló épüljön ki, mely a barangoláson túl a módszeres kutatómunkát is jól kiszolgálja. Mivel a projekt hosszú távú célja az, hogy a portál az időközben esetleg módosuló struktúrájú kutatáson túlmutatva a magyar színháztörténet digitális tudásközpontjává váljon, az OSZMI a *szinhaztortenet.hu* doménnevet regisztrálta a portál elérésére.

A honlap egyediségét részben az biztosítja, hogy a publikált szövegek mellett a hivatkozott és kapcsolódó források is közvetlenül elérhetők. Ezek egyrészt segítik a mélyebb, megértőbb olvasást, másrészt döntően vizuális tartalmakkal teszik érzékletessé az olvasottakat, bátorítva a továbbgondolást, az újabb kapcsolatok felállítását.

A színháztörténet vagy a kortárs színház kutatóinak az összetett keresés nyújtja a legnagyobb segítséget. Az OSZMI célja az, hogy az új adatbázis gerincét adó Színházi Adattárban az összes magyarországi és magyar nyelvű hivatásos színházi előadás részletes adatai szerepeljenek, beleértve a részletes szereposztást és ennek esetleges változásait. Gyakori kutatói igény egy-egy színész szerepkatalógusának elkészítése vagy egy adott évad új bemutatóinak kilistázása. Összetettebb feladat, ha egy adott rendező vagy társulat előadásairól szóló összes kritikát vagy fényképet kívánjuk lekeresni, hiszen ekkor a rendszernek az adott cikk vagy fényképrekordokhoz kapcsolódó előadásrekordok adataiban kell keresnie. Nem

egyszerű kihívás megjelölni azokat az adatmezőket, amelyeket kereshetővé kívánunk tenni, úgy hogy a lista mégse legyen áttekinthetetlenül hosszú a felhasználó számára. Egy-egy konkrét kutatói igény néha az eredeti prioritások felülvizsgálására készlet: például igény mutatkozott arra, hogy egy adott művész által elnyert díjakra is lehessen keresni. Igen hasznos lehetőség, hogy az esetlegesen túl tág találati listát facettákkal tovább lehet szűrni például alkotó, kor, gyűjtemény vagy dokumentumtípus alapján. A színházi terület komplexitását mutatja az is, hogy néha igen nehéz értelmezni és alkalmazni az akár nagyon szabványos adatmezőket is. Ebben a rendszerben például a cím vonatkozhat egy drámai mű vagy egy előadás címére, de egészen eltérő egy interjúnak vagy kritikának esetleg kéziratos levélnek a címe, míg bizonyos források, mint a jelmezek, díszlettervek vagy fényképek esetén, címről nem beszélhetünk. A szerző kérdése az előadó-művészetben még ennél is összetettebb. Ahhoz, hogy a keresések megbízható, teljes és nem redundáns eredményeket adjanak, elengedhetetlen a szótármezők, besorolási állományok és névterek megtisztítása, amely ismertén időigényes feladat.

Jövőbeni tervek

Az adatbázis fejlesztése természetesen nem állhat meg. A további adatrögzítés, a digitalizáltság teljessé tétele maga is óriási feladat. Az adatgazdagítás fentebb már említett kézi, sőt esetleg gépi lépései is előttünk állnak. Idővel sor kerülhet a művek és a szerepek besorolási rekorddá való alakítására, ez utóbbiak esetében inkább csak részben. A szerepek között ugyanis egyszerű keresztnévek (Erzsi, Anna, János), illetve foglalkozásnevek, vagy egyéb köznevek, katona, pincér, anya nagy számban vannak jelen.

A GDPR-elvek érvényesítése még az eddigi gyakorlaton túlmenően is egy külön további feladat. Érdekes módon, egy ilyen kortárs személyeket is tartalmazó adatbázisban, nemcsak a könyvtárakra jellemző feladatot, az olvasók adatainak megfelelő kezelését kell szem előtt tartani, hanem a színészek és egyéb közreműködők személyes adatait is körültekintéssel kell kezelni és védeni. A publikálásra nem került információkat a személyeknél csak engedélyükkel, vagy hozzátartozóik engedélyével lehet az adatbázisban nyilvánosságra hozhatóként megjelölni. Amennyiben az adatok nem nyilvánosak, azok védelméről a rendszer természetesen az egyéb érzékeny adatokkal együtt gondoskodik. Az intézmény munkatársainak is

lehetnek az adatbázisban személyes adatai. Ezekről ugyancsak gondoskodni kell, de ha az ilyenek az intézményben végzett munkájukkal összefüggenek, már nemcsak hogy rögzíthetők, de adott esetben nyilvánosságra is hozhatók. Például olyan esetben, amikor valaki gyűjtő, kutatómunkát végez valahol.

Összefoglalás

Szűk másfél év leforgása alatt a *Színházi emlékezet* projekt jelentős eredményeket mutathat fel. Sikerült letenni egy átfogó magyar színháztörténeti kutatás alapjait, mely a múltbéli előadások multimediális nyomait vizsgálatának középpontjába helyezi. A rövid idő alatt a kis kutatócsoportnak számos különálló feladatot kellett megvalósítani úgy, hogy közben a nagyközönség számára is értékelhető eredményt tudjon felmutatni. A színháztörténeti kutatás módszertanának kialakítása, a dokumentáció begyűjtése, a digitalizáció, az interjúk elkészítése és szerkesztése, a tanulmányok megírása mellett az új adatbázis kialakítása, a régi adatok migrációja és a projekt portáljának felépítése mind egyidőben kellett, hogy megvalósuljon. Ugyanakkor a múzeum munkatársai bíznak benne, hogy ez az alapozó munka lehetővé teszi, hogy a jövőben térben és időben is könnyen kiterjesztheti a kutatás irányát és apránként minden magyar nyelvű színház a program részévé válik. Az eddig elvégzett „próbaúrások” igazolni látszanak a korszerű színháztörténeti kutatások irányainak a projekt során felvetett hipotéziseit: az integrált adatbázisok, tudásrendszerek létrehozásának fontosságát, a helyi múzeumokkal, levéltárakkal való szoros együttműködés és adatmegosztás lehetőségeit, a tágabb szociológiai, társadalomtörténeti és intézménytörténeti kontextus elemzésének jelentőségét, a nagyméretű, jól metaadatolt digitális forrásgyűjtemények új perspektíváit.

Hivatkozások

- ¹ Molnár Klára: Az intézet számítógépes adatbázisának fejlődése, Gajdó Tamás (szerk.): Digitális
- ² BÁNKI Zsolt, Mészáros Tibor, Németh Márton, Simon András: Azonos személyekre vonatkozó név besorolási rekordok automatikus felderítése a PIM adatbázisában = Tudományos és Műszaki Tájékoztatás, 63. évfolyam (2016) 12. szám
- ³ Rostás Péter: Várda a vártán. In: Budapest 2004/6 (Augusztus 26-28.)
Deák Zoltán: A gellérthegyi Citadella. In Várak 2016

- ⁴ Knowledge organization systems (KOS) Article in Knowledge Organization · January 2008 Marcia Lei Zeng
- ⁵ BOGNÁR Katalin [2010]: *Tudásalapú rendszerek és technológiák*.
- ⁶ Fülöp Endre: A szemantikus háló két fogalma, a katalógusok új generációja és a könyvtárak szerepe. In: TMT, [65. évf. 7-8. sz. \(2018\)](#)
- ⁷ LIDO: Lightweight Information Describing Objects, eredetileg múzeumi adatcsere formátum, ezen alapul például az EDM (European Data Model) is, mely az Europeanának felküldendő dokumentumok metadatumainak adatszerkezetét határozza meg. Egyes múzeumi rendszerek a LIDO továbbfejlesztésével adatcsere formátum mellett már belső adattárolási formátum funkciójával is felruházzák. A Monguz KFT által fejlesztett Huntéka-M múzeumi gyűjteménykezelő rendszer LIDO inputot konverzió során fogad, LIDO formátumot adatcsere formátumként exportál, illetve a felhasználói felületen megjelenít, de (nem elegendő mértékű tagoltsága miatt) belső formátum céljaira nem alkalmazza. információ a LIDO formátumról: <http://www.lido-schema.org/schema/v1.0/lido-v1.0-specification.pdf>
- ⁸ BÁNKI Zsolt, Mészáros Tibor, Németh Márton, Simon András: Azonos személyekre vonatkozó név besorolási rekordok automatikus felderítése a PIM adatbázisában = Tudományos és Műszaki Tájékoztatás, 63. évfolyam (2016) 12. szám
- ⁹ Lengyel Monika, Szabó Attila: Színházi Emlékezet portál / Konferencia előadás: Muzeum@digit 2018 Budapest
- ¹⁰ Az ábrát Simon András az adatbázisból kinyert adatok alapján készítette.
- ¹¹ Lengyel Monika, Szabó Attila: Színházi Emlékezet portál / Konferencia előadás: Muzeum@digit 2018 Budapest
- ¹² Lengyel Monika, Szabó Attila: Színházi Emlékezet portál / Konferencia előadás: Muzeum@digit 2018 Budapest
- ¹³ A múzeumi gyakorlatban az egyes műtárgy típusoknak szakleltárkönyvek felelnek meg.
- ¹⁴ Múzeumi gépi leltározás esetében ez a 20/2002. (X.4.) NKÖM rendelet.
- ¹⁵ A szótárelemeket a cikk későbbi részében felsoroljuk.
- ¹⁶ <https://www.theatre-architecture.eu/hu/db.htm>
- ¹⁷ Simon András: A Petőfi Irodalmi Múzeum személynévterének publikálása a szemantikus weben = Networkshop 2016 Országos konferencia. IIF program. Debrecen, (2016)

- ¹⁸ Mindhárom weboldalon, hasznos és értékes információk sokasága van feltüntetve, de távolról sem mindig azonos szerkezetben, és válogatásban. A PIM-es és az OSZMI-s rekordról, e sorok szerzői, mint illetékesek közlik az olvasóval, hogy nincs minden információ a webes megjelenítő felületen nyilvánosságra hozva. Pl. Az ismert művész születési neve, Grósz Dezső fel van tüntetve az OSZMI adatbázisban, de ezt az információt a webes felületen a múzeum nem jeleníti meg. E cikk egyik szerzőjének, Simon Andrásnak szilárd meggyőződése, hogy a külső adatforrások egyelőre nem tudják helyettesíteni egy gyűjteménykezelő rendszer saját adatbázisában tárolt besorolási rekordjait, és ennek a lábjegyzetben közölt linkek is ékes bizonyossággal szolgálnak.

Irodalom

BÁNKI Zsolt, Mészáros Tibor, Németh Márton, Simon András: Azonos személyekre vonatkozó név besorolási rekordok automatikus felderítése a PIM adatbázisában = TMT, 63. évfolyam (2016) 12. szám

BOGNÁR Katalin [2010]: *Tudásalapú rendszerek és technológiák*.

Fülöp Endre: A szemantikus háló két fogalma, a katalógusok új generációja és a könyvtárak szerepe, TMT 65. évf. 7-8. sz. (2018)

Knowledge organization systems (KOS) Article in Knowledge Organization January 2008 Marcia Lei Zeng

Rostás Péter: Várda a vártán In: Budapest 2004/6 Augusztus 26-28.)

Molnár Klára: Az intézet számítógépes adatbázisának fejlődése, Gajdó Tamás (szerk.): Digitális Színháztörténet, Színháztudományi Szemle 38, OSZMI, 2009, 137.

Beérkezett: 2019. V. 13-án.



Simon András

ügyfélmenedzser,
ügyfélkapcsolati, projektek vezetője
Monguz Kft.
E-mail: asimon@monguz.hu



Szabó Attila

a PIM-OSZMI igazgatóhelyettese.
E-mail: szabo.attila@oszmi.hu