

Berze Lajos

## **SZIA SZIE – repozitóriumi gyűjtemény kialakítása a SZIE Kosáry Domokos Könyvtár és Levéltárban**

*Mára a felsőoktatási intézményeknek kiemelt, előírt feladatává vált, hogy az oktató- és a kutatómunka eredményeképpen létrejövő szellemi vagyont rendszeresen, visszakereshetően megőrizzék, s az így összegyűjtött, digitalizált formában rendelkezésre álló tudást a lehető legszélesebb kör számára elérhetővé tegyék, szolgáltatassák. A Szent István Egyetemen a Kosáry Domokos Könyvtár és Levéltár (SZIE KDKL) is vállalta e nemes küldetés megvalósítását. Írásomban ismertetem a Szent István Archívum (SZIA) kialakításának folyamatát, valamint a PhD gyűjtemény példáján keresztül azokat a fejlesztési feladatokat, amelyekkel munkánk során találkozunk.*

### **A repozitórium rövid története**

A Kosáry Domokos Könyvtár és Levéltárban a Szent István Egyetem szellemi vagyonához, örökségéhez tartozó digitális dokumentumok szervezett gyűjtése és közzététele 2008–2009-ben a Magyar Digitális Képkönyvtár (MDK) projektben való részvétellel kezdődött meg. Ekkor alakították ki azt a JaDoX alapú gyűjteményt, amely alkalmas volt a digitalizált képanyagot metaadatokkal ellátva, szabványos adatátviteli protokollok segítségével az MDK számára „átadni”. Jelenleg a képkönyvtári gyűjtemény több mint 1100 rekordot tartalmaz, amelyből több mint 500 található meg az MDK-ban (1. ábra).

A munka 2011-ben folytatódott, amikor is a „Tudásdepó – Tudásvásár” címet viselő TÁMOP pályázat keretében könyvtárunk az egyetem doktori iskoláiban írt és megvédett PhD disszertációk elektronikus feldolgozását és közzétételét vállalta. A pályázati időszak során, egy év alatt, majd 230 PhD disszertáció, s a hozzájuk tartozó magyar és angol nyelvű tézis – azaz összesen közel 900 dokumentum – került betöltésre és formai, tartalmi feltárára (metaadatolásra). A gyűjtemény folyamatosan bővül, jelenleg 435 teljes disszertáció található az állományban (2. ábra).



1. ábra Az „Atlas der Alpenflora” című könyv képei a Képkönyvtárban

PhD SZENT ISTVÁN EGYETEM PhD

Közkezdőoldal Keresés **Böngészés** Kapcsolat Segítség Oktatóanyag Statistika Doktor iskolák SZIA

Közkezdőoldal >> Böngészés

Szűrőfeltételek: **Cím**

Partnerek: **Összes**

Böngészés

Sorok száma: **435**  
Rekordok száma: **435**

Rekordok száma: **10 db/oldal**

1. oldal

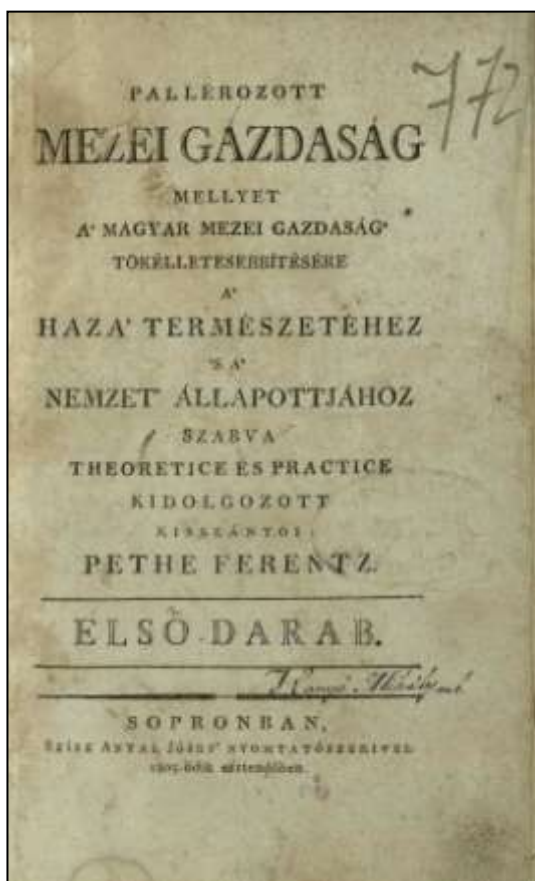
- A 16-37ág szabályozó régió represszorok és operátorok - 1 db
- A 2007-ben kibontott pénzügyi válság hatásainak értékelése a befektetési alapoknál Magyarországon - 1 db
- A biodiverzitás értékelése szilárdanyag közgazdasági megközelítésből - 1 db
- A biotechnológiai módszereken alapuló növénytermesztés két új eszköze: a stáka proteáz inhibitor és pPROGMO transzformációs vektor - 1 db
- A biocentrizmusok felhasználásának pénzügyi vonatkozásai, a szabályozás továbbfejlesztése - 1 db
- A Borsógyar tápellékes a növénytermesztés szerepeiről - 1 db**
- A bor eloszlásos csatornáinak marketing szempontú értékelése a borvásárlási szokások és a termelési magatartás vizsgálatán keresztül - 1 db
- A borfogyasztói és -visszairó magatartás fontosabb összefüggéseinek vizsgálata - 1 db
- A burgonya Y vírus köpenyfehérje gén által indukált rezisztencia kialakítása burgonya- és dohányfajtákban, és ezek szántóföldi értékelése - 1 db
- A Campylobacter előfordulása az élelmiszerláncban, valamint környezeti tényezők hatása túlmérsékre és pszichikusokra - 1 db

2. ábra Találati halmaz a PhD gyűjteményben

A 2012–2013-ban lezajlott sikeres NKA-pályázatnak köszönhetően könyvtárunk folytathatta az évekkorábban már megkezdett, muzeális értékű könyvek, folyóiratok digitalizálását célzó programját. A pályázat keretében Nagyváthy Já-

nos és Pethe Ferenc 1850 előtt megjelent, agrár vonatkozású művei közül 8 könyvet, összesen 5600 oldalt digitalizáltak és helyeztek el a kialakított repozitóriumi gyűjteményben (3. ábra).





3. ábra Régi Ritka Könyvek a Szent István Archívumban

2014-ben kezdődhetett meg az egyetemi oktatók, kutatók publikációinak feldolgozása, gyűjteménybe szervezése. Ez idáig mintegy 400 folyóirat- és konferenciacikket tettünk hozzáférhetővé az érdeklődők számára (4. ábra).

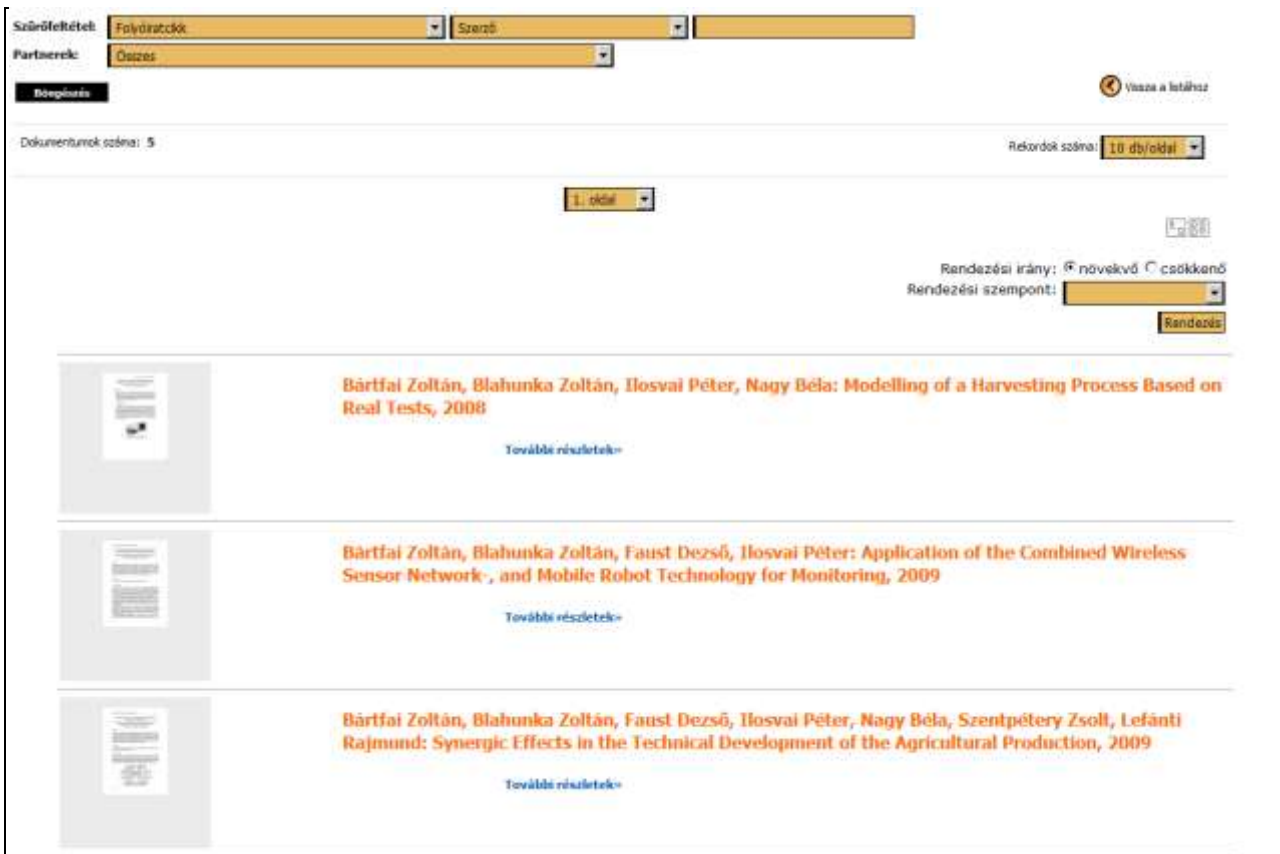
Ezek a gyűjtemények képezik az egyetem intézményi repozitóriumának, a Szent István Archívumának a törzsét.

### A PhD gyűjtemény kialakítása

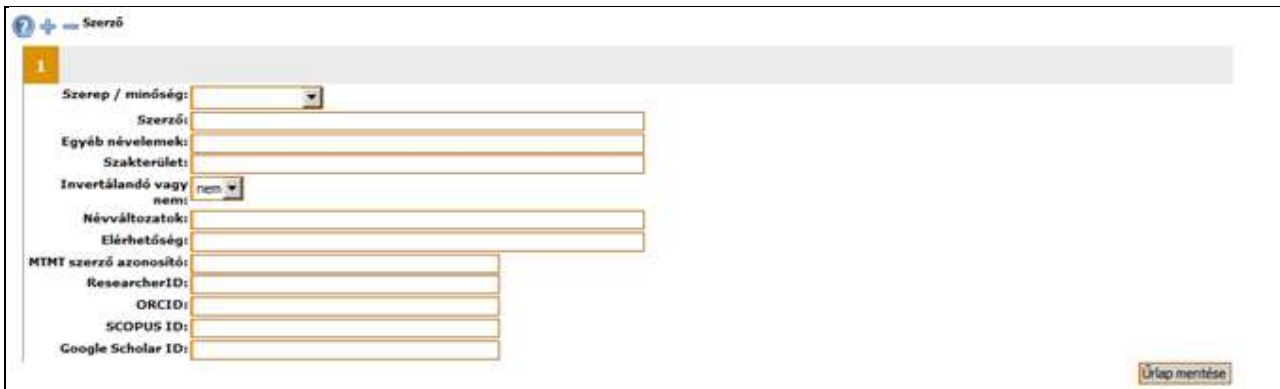
A repozitórium valamennyi gyűjteménye esetében alapelv volt, hogy szerkezeti kialakításuk a lehető leginkább „kompatibilis” legyen a *Magyar Tudományos Művek Tárával*, azaz a közzétett publikációkról az MTMT-ben nyilvántartott valamennyi információt tárolni tudjuk a repozitóriumban is. Természetesen az egyes dokumentumokról több olyan információt is rögzíteni kívántunk, amelyeket az MTMT-ben nem rögzítenek. Az ilyen információk köréről, a mezők megnevezéséről, a kitöltés mód-

járól, illetve a mezők úrlapon elfoglalt helyéről a könyvtár munkatársainak részletes egyeztetéseit követően született döntés.

Mind a tervezés, mind a kivitelezés során figyelembe kellett venni, hogy a JaDoX nem támogatja az olyan típusú egységesített, segédtablákat alkalmazó adatfelvitelt, mint amilyen segítségével az MTMT-ben az információk egy részének rögzítése történik (pl. szerzők, szerkesztők nevének vagy a folyóiratok címének esetében). Így a gyűjtemény adatsémáját és adatbeviteli úrlapját úgy kellett kialakítanunk, hogy az ilyen adatsoportok valamennyi elemét minden rekord esetén rögzítésék, s ahol ez szükséges (pl. több szerző, több közreműködő esetében), ott az adott mezőcsoport többszörözhető legyen. Amint az az 5. ábrán látható, az ilyen adatsoportok esetében egy plusz-jel látható, amelyre adatrögzítés közben kattintva a teljes adatsoport többszörözhető. (A szükségtelenül megismételt adatsoportokat a mínusz jelre kattintva lehet törölni.)



4. ábra Egy szerző művei az egyetemi oktatók, kutatók publikációinak gyűjteményében



5. ábra A szerző(k) adatainak rögzítésére szolgáló, többszörözhető mezőcsoport

Az adatfelvitel ilyen módja természetesen okoz némi többletmunkát, azonban csak így volt biztosítható, hogy a szerzőkre, szerkesztőkre, folyóiratokra vonatkozó valamennyi adatot rögzítsék.

A JaDoX szoftver egyik hatalmas előnye, hogy az adatfelvitel során a teljes szövegű állományok leíró adatai nemcsak manuálisan rögzíthetők, de szab-

ványos MARC formátumú állományok segítségével is importálhatók az adatbázisba.

Mivel a projekt során feltöltésre kerülő PhD disszertációk mindegyike fel volt dolgozva már könyvtárunk integrált rendszerében, így – a HunTékából való exportálást követően – a szükséges MARC állományok is rendelkezésünkre álltak. Ebből adó-

dóan a disszertációk betöltése a repozitóriumba – a megfelelő megfeleltetéseket követően – nagymértékben automatizálható volt. A megfeleltetés céljára tökéletesen alkalmasnak bizonyult egy Excel tábla, amelyben az adatbázis egyes mezőinek esetében rögzítettük, hogy oda mely MARC mező tartalma kerüljön be az importálás során.

Az ekként összerendelt elemek technikai, adatbázisban való megfeleltetését a *Monguz Kft.* munkatársai végezték el. A megfeleltetések működését tesztfeltöltésekkel ellenőriztük, így az esetleges hibák már a tömeges adatfelvitel előtt kideríthetők és javíthatók voltak (6. ábra).

A	B
Cím	
Cím:	245 \$a
Alcím:	245 \$b
Párhuzamos cím:	741 \$a \$b
Besorolási cím:	245 \$a-ból az első // és a második // közötti rész
Közös (összefoglaló) cím	
Rész cím	
Eredeti cím	240 \$a \$b; 130 \$a \$b; 730 \$a \$b
Szerző	
Sorszám:	
Szerep / minőség:	
Besorolási név:	100 \$a
Utónév:	100 \$j
Egyéb névelemek:	100 \$b \$c \$d \$g \$m
Invertálható vagy nem:	
Névváltozatok	
Elérhetőség	
Doktori iskola	
Sorszám:	
Szerep / minőség:	
Doktori iskola:	710 \$a, \$b, 110 \$a \$b
Székhely:	710 \$c, 110 \$c
Országkód	
Elérhetőség	
Beviteli dátum	
Beviteli dátum:	
Dátumra vonatkozó megjegyzés	
Megjelenési dátum	
Megjelenési év:	260 \$c
Dátumra vonatkozó megjegyzés	

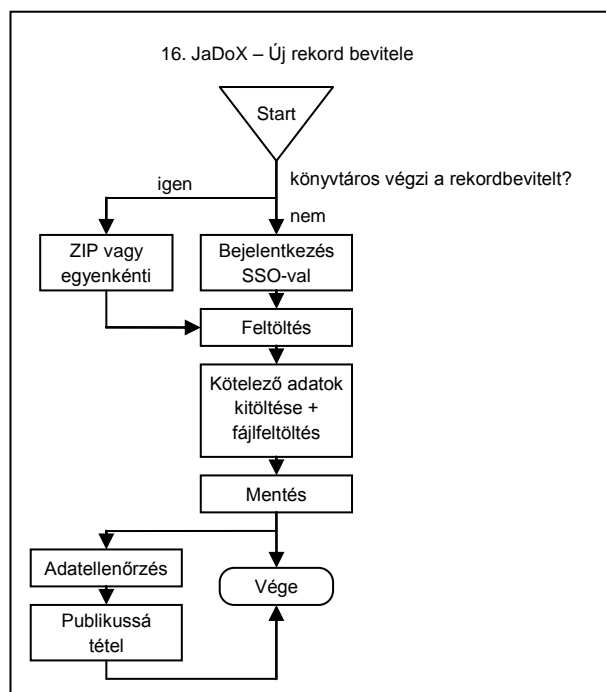
6. ábra Sémamegfeleltetés

A sémamegfeleltetést követően került kidolgozásra a feltöltési folyamat (7. ábra), amelynél elemekre bontva meghatároztuk az egyes folyamatlemlékek végrehajtásához szükséges időigényt is (8. ábra).

A feltöltésre kiválasztott disszertációkat a JaDoX rendszer „Tömeges feltöltés” funkciójának segítségével importáltuk az adatbázisba. A technológia lényege, hogy az összetartozó dokumentumokat (teljes szövegű állományok, illetve a metaadatokat tartalmazó MARC állományok) ZIP állományba tömörítve, egy lépésben lehet az adatbázisba töl-

teni. (Egy tömörített állomány több dokumentum összetartozó fájljait is tartalmazhatja.)

Ahhoz, hogy az összetartozó dokumentumok (disszertációk esetében az értekezés teljes szövegét, valamint a magyar és angol nyelvű téziseket és a leíró adatokat tartalmazó MARC állomány) a rendszer számára „felismerhetőek legyenek, három számjegyű sorszámozást kell alkalmazni a dokumentumfájlok nevében. A sorszámozás 001-től indulhat. További megkötés, hogy a fájl nevének illeszkednie kell az alábbi reguláris kifejezésre: [az09\_]+(\_[09][09][09])?“.kiterjesztés”. Vagyis a név kisbetűkből, számokból és '\_'ből kell, hogy álljon, amelyet ha van sorszámozás, egy '\_'-nak kell követnie és a háromjegyű sorszámnak, majd a fájl kiterjesztésének: például valami23.jpg, konyv\_002.tiff, valamilyen\_dokumentum.xml.



7. ábra A feltöltési folyamat sémája

Egy példával megvilágítva a több fájlból álló dokumentumok nevezékét (9. ábra):  
 doktoriertekezés.MRC – MARC állomány,  
 doktoriertekezés\_0001.PDF – disszertáció,  
 doktoriertekezés\_0002.PDF – magyar tézis,  
 doktoriertekezés\_0003.PDF – angol tézis.

Szükséges idő:  
A PhD dolgozatok egyenkénti feltöltésének időzítése:<sup>5</sup>

	Tevékenység megnevezése	Tevékenység időtartama	Tevékenység iránya
Előkészítés	Dokumentumok kiválasztása	kb. 8 óra	A feldolgozandó dokumentumok körének kiválasztása
	Dokumentumok beszerzése	kb. 8 óra	A feldolgozandó dokumentumok körének beszerzése
	Dokumentum kiválasztása	1 perc	Az aktuálisan feltöltendő dokumentum kiválasztása
	Dokumentumtípus kiválasztása	1 perc	A feltöltendő dokumentum dokumentumtípusának meghatározása
	Metaadat-típus kiválasztása	1 perc	A feltöltendő dokumentumhoz tartozó metaadat-típus meghatározása
Feltöltés	Dokumentum feltöltése	3 perc	A kiválasztott dokumentum betöltése a rendszerbe
	Feltöltés ellenőrzése	5 perc	Ellenőrzés a feltöltés sikerességét illetően
	Metaadatok szerkesztése	18 perc	A feltöltött dokumentum metaadatainak szerkesztése
Ellenőrzés/javítás	Metaadatok ellenőrzése	5 perc	A feltöltött metaadatok helyességének ellenőrzése
	Javítás	5 perc	Az esetleges hibák javítása
Összesen (a munkafolyamat ismétlődő elemeit figyelembe véve):		39 perc/dokumentum	

8. ábra Az egyes folyamatok időigényének meghatározása

Tömeges fájlfeltöltés    Metaadat feltöltés    Doktori iskolák    SZIA

Template >> Speciális rekord adminisztráció >> Tömeges fájlfeltöltés

### ZIP tömörített állomány feltöltése

**Feltöltéshez szükséges adatok:**

Könyvtár:

Marc esetén típus:

Marc esetén kódolás:

ZIP fájl:

A fájlok feltöltése folyamatban...

Kérem várjon türelemmel!

9. ábra A tömeges fájlfeltöltés képernyője

A betöltött ZIP állományt a rendszer „kibontja”, s a fájlnevek alapján összetartozó állományokból (jelen esetben PDF-ekből és HUMARC állományból) összeállítja az adott dokumentum repozitóriumi rekordját (10. ábra).

Mivel a TÁMOP pályázat keretében nemcsak a gyűjtemény kialakítását vállaltuk, hanem azt is, hogy a gyűjtemény a saját szerveren kívül a *Budapesti Corvinus Egyetem* mint konzorciumvezető által működtetett, a konzorciumi gyűjtemények tekintetében aggregátorként funkcionáló portálján is (<http://tudasdepo.uni-corvinus.hu/>), így a repozitórium „mellé” telepítésre került a metaadatok aratását szolgáló OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*) protokoll is (11. ábra). Ennek teszteléséhez, ellenőrzéséhez az interneten elérhető, az MTMT munkatársai által is ajánlott validátorokat használtuk.

Noha a pályázati projekt időközben lezárult, a repozitórium, s ezen belül a PhD gyűjtemény technikai és tartalmi fejlesztése nem állt meg. A további fejlesztéseket a nemzeti felsőoktatásról szóló 2011. évi CCIV. törvény 53/A. §-ban foglalt előírá-

sok tették szükségessé. Ezeknek megfelelően egyetemünknek is meg kellett kezdenie a doktori értekezések DOI azonosítóval történő ellátását.

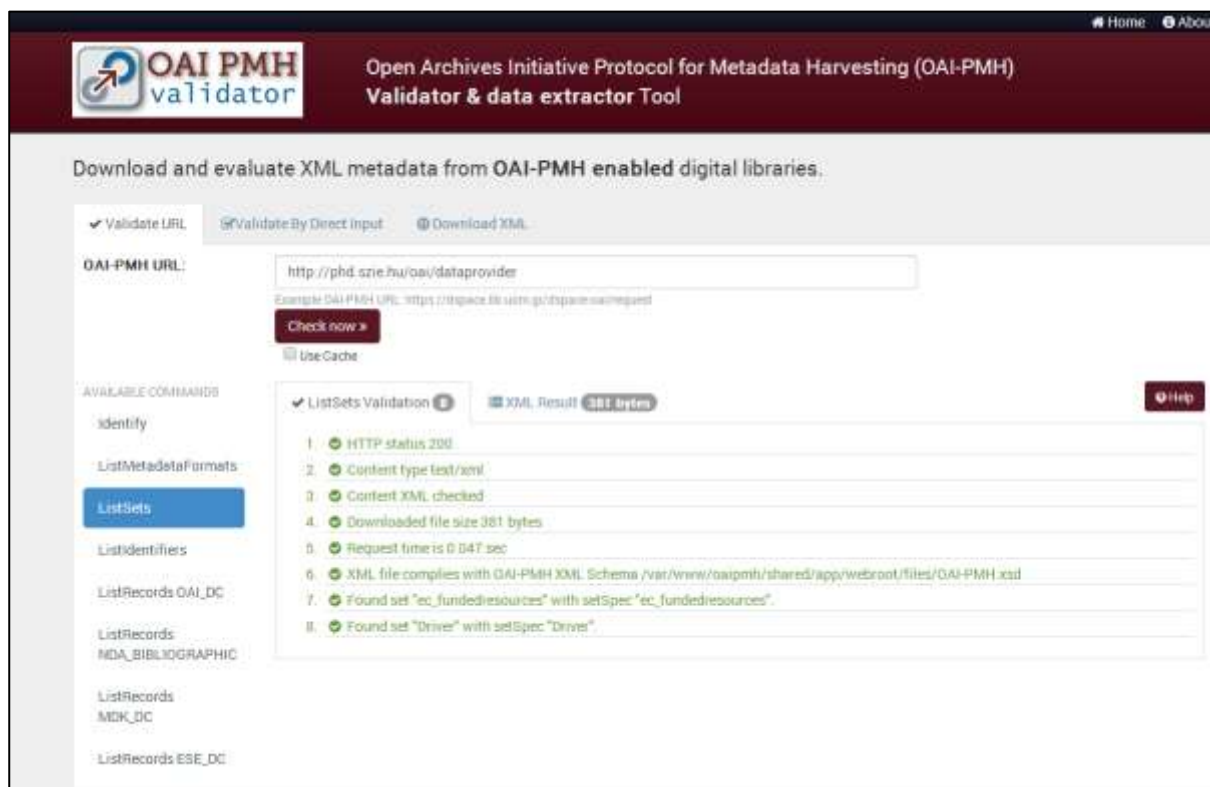
A fejlesztéseket az MTMT Repozitórium Minősítő Bizottság által kiadott protokollban megfogalmazott előírásokhoz, igényekhez igazodva kezdtük meg. (Az MTMT Repozitórium Minősítő Bizottság 2014. április 23-i ülésén ideiglenes minősítést ítélt meg a Szent István Egyetem repozitóriuma számára.) Technikai oldalról a legnagyobb feladatot a SWORD (Simple Web-service Offering Repository Deposit) protokoll telepítése, illetve a JaDoX adatbázishoz való illesztése jelentette.

A gyűjtemény tartalmi fejlesztése a pályázati projekt megkezdése óta folyamatos. Jelenleg az érdeklődők 435 doktori értekezést tekinthetnek meg a Szent István Archívumban. Annak érdekében, hogy a gyűjteményhez a lehető legszélesebb körben biztosíthassunk hozzáférést, repozitóriumunkat bejegyeztettük a legismertebb regiszterekbe (ROAR, OpenDOAR).

The screenshot displays a record for a PhD dissertation. On the left, there is a PDF viewer showing the cover page of the dissertation, which includes the Szent István Egyetem logo and the title 'VÁLLALATI KAPCSOLATI RENDSZEREK A HAZAI KISKERESKEDELMI ÖVAKORLATBAN'. Below the viewer, there are links to download the dissertation in PDF format. On the right, a metadata sidebar provides detailed information about the record, such as the title, author, subject, and publication date.

Cím:	Vállalati kapcsolati rendszerek a hazai kiskereskedelmi övakerlatban
Alcím:	A hosszú távú üzleti kapcsolatok vizsgálata
Részcím:	doktori értekezés
Szerep / minőség:	szerző
Szerző:	Király Éva
Szakterület:	gazdálkodás és szervezéstudományok
Invertálendő vagy nem:	nem
Szerep / minőség:	közreadó
Doktori iskola:	Szent István Egyetem, Gazdálkodás és Szervezéstudományi Doktori Iskola
Doktori iskola székhelye:	Gödöllő
Ország:	Magyarország
Beviteli dátum:	2014-03-05
Megjelenési év:	2011
Nyelv:	magyar
A típus neve:	disszertáció
Témakör:	Közgazdaságtan, gazdaság
Altémakör:	Vállalatok
Tárgyszó:	vállalati kapcsolatok szervezetek kiskereskedelem üzleti kapcsolatok Magyarország PDF dokumentum
A formátum neve:	feldolgozás alatt
Az adatrekord státusza:	ellenőrzött
A dokumentum státusza:	ellenőrzött
Kiadó neve:	Szent István Egyetem

10. ábra Egy PhD disszertáció „borítóoldala” a Szent István Archívumban



11. ábra Az OAI-PMH Validator & data extractor (validator.oaipmh.com) ellenőrző képernyője

Mint látható, az elmúlt néhány év során könyvtárunk is nagy lépéseket tett az egyetemi szellemi vagyon széles körű hozzáférhetőségének biztosítása érdekében. Feltérképeztük, beazonosítottuk a digitálisan rendelkezésre álló, vagy digitalizálható tudásvagyon elemeit, és megteremtettük azt az informatikai infrastruktúrát, amely révén megkezdődhetett a feldolgozás és a szolgáltatás. E lépésekkel egy olyan úton indultunk el, amelyen nincs, nem lehet megállás. Az elkövetkezendő időszak feladata, hogy valamennyi vagyonelem teljes körű

gyűjtése, feldolgozása és szolgáltatása megvalósuljon.

Beérkezett: 2014. XI. 5-én.

