

tóknak arra, hogy hogyan lehet korszerűsíteni a rendszereiket, beépítve hasznos webkettes megoldásokat, vagy éppen olyan régi, web előtti funkciókat, amelyeket korábban csak az információkereső szakemberek ismertek és használtak, mivel azok eddig nem voltak elég felhasználóbarátok.

/JACSÓ Péter: Traditional scholarly publishers and Web 2.0: the case of Springer. = Online Information Review, 35. köt. 2. sz. 2011. p. 301–315./

(Drótos László)

Intézményi repozitórium fejlesztése a DigiTool rendszerrel

Bevezetés

2006 végén, a *Colorado State University* stratégiájának részeként az egyetemi könyvtár (a továbbiakban: CSUL) megbízást kapott egy intézményi szintű repozitórium kialakítására, mely amellett, hogy lehetővé teszi a nyílt hozzáférést a tudományos eredményekhez, az egyetem szellemi termékeinek valamint történeti dokumentumainak központi őrzőhelyeként is funkcionál. A CSUL szakemberei több szóba jöhető szoftvert is megvizsgáltak, és a DSpace, a Digital Commons, a DigiTool, a Fedora, a Hive, illetve a Symposia közül végül a DigiTool-ra, az *Ex Libris* cég digitális állománykezelő rendszerére esett a választásuk. A 2007 májusában beszerzett rendszer mellett több érv is szól: egy viszonylag kiértelt és a forgalmazó által támogatott termék, mely kezeli az elterjedt fájlformátumokat és metaadatszabványokat, emellett kellemes felhasználói felülete van, olyan fontos funkciókkal, mint például beépített JPEG2000 nézegető és METS megjelenítő, illetve a teljes szövegben való keresés és a felhasználók általi dokumentumbeadás lehetősége. Mivel a könyvtár már korábban is használt *Ex Libris* termékeket (SFX és *MetaLib*), lényeges szempont volt az is, hogy a DigiTool könnyen integrálható ezekkel. A *digitooll.library.colostate.edu* címen található repozitórium hivatalos megnyitója 2008 márciusában volt, vagyis szűk egy év állt rendelkezésre arra, hogy könyvtári és informatikai szakemberek telepítsék, beállítsák és teszteljék a rendszert. 2009 decemberében már csaknem 15 ezer digitális objektum volt a repozitóriumban, köztük tézisek és disszertációk, tanszéki dolgozók publikációi, hallgatók kutatási prezentációi, konferenciaanyagok, valamint levéltári dokumentumok és képek. A könyvtár korábban, 2001 és 2007 között egy *CONTENTdm* nevű rendszerrel kezelte a digitális gyűjteményeit (<http://www.digital.library.colostate.edu>). Ezek közül két kisebbet már átköltöztettek a DigiTool alá és várhatóan továbbiakat is migrální fognak az új rendszerbe.

E cikkben a szerzők a DigiTool rendszerrel szerzett tapasztalataikat foglalják össze és részletesen bemutatják azokat a konfigurációs beállításokat, saját kiegészítéseket és integrációs megoldásokat, amelyekkel sikerült egy, az elvárásoknak megfelelő intézményi repozitóriumot kialakítaniuk. Eddig még csak néhány publikáció jelent meg erről a témáról. Közülük említést érdemel a *Liverpool John Moores University* munkatársainak cikke [1], akik 2005-ben az elsők között tesztelték a DigiTool-t ilyen szempontból, és úgy találták, hogy az alkalmas egy egyetemi digitális archívum működtetésére, előnyeként pedig a *Ex Libris* termék támogatását és a sokféle dokumentumtípus kezelésének képességét emelték ki az ingyenes alternatívákkal szemben. A kanadai *McGill University* szintén a DigiTool rendszert választotta az OAI-kompatibilis elektronikus szakdolgozat-gyűjteményéhez; az erről szóló tanulmány [2] azonban inkább a munkafolyamatot mutatja be, nem tér ki a DigiTool tesztelésével kapcsolatos tapasztalatokra.

A DigiTool képességei

A DigiTool rendszer különféle adminisztrátori és felhasználói funkciói hat különböző webes alkalmazás segítségével érhetők el:

- A *Resource Discovery* nevű modul a végfelhasználók számára készített felület, amellyel hozzáférhetnek a gyűjteményekhez és a digitális dokumentumokhoz.
- A *Management* modullal végezhető a rendszergazdai feladatok, például: karbantartás, konfigurálás, beadási űrlapok szerkesztése, metaadatmezők hozzáadása vagy módosítása, stabil azonosítók generálása és közzététele, jelentések készítése.
- A *Collection Management* alkalmazás a virtuális gyűjtemények létrehozását és kezelését segíti.
- A felhasználók a saját anyagaikat a *Deposit* modulon át tudják beküldeni.

- A könyvtárosok az *Approver* modullal tudják ezeket megnézni és jóváhagyni vagy visszautasítani.
- A *Web Ingest* nevű eszközzel maguk is betölthetnek digitális objektumokat és metaadatokat a DigiTool adatbázisaiba.

Mindezek mellett van még egy Windows gépeken önállóan futtatható *Meditor* (Metadata Editor) elnevezésű program is, mellyel szintén elérhetőek a fent említett adminisztrátori funkciók és néhány további eszközt is nyújt a metaadatok és a digitális objektumok kezeléséhez.

A DigiTool rendszer „digitális entitás”-okat tárol és kezel. Egy ilyen entitás több alkotóelemből áll: egy állandó azonosító (PID), egy vezérlő rész különféle jellemzőkkel (címké, megjegyzés, entitás típus, használati típus stb.), többféle metaadattípus (adminisztrációs, leíró, technikai, megőrzési), kapcsolat más digitális entításokkal (pl. „Megjelenési formája”, „Tartalmazza”, „Része”), valamint a digitális tartalmat hordozó adatfolyam(ok). A „használati típus” négyféle lehet: ARCHIVE, VIEW, THUMBNAİL és INDEX. Egy digitális objektum egy vagy több ilyen entitásból állhat. Például egy PDF-formátumú szakdolgozat esetében tipikusan három entitás van: maga a PDF fájl (VIEW típus), egy címlapkép JPG formátumban (THUMBNAİL típus) és a dokumentum szövege HTML-ben (INDEX típus). Az ilyen egyszerű objektumok mellett a DigiTool saját belső struktúrával rendelkező komplex objektumokat is képes kezelni (mint amilyen pl. egy folyóiratszám). Ugyancsak nagyon rugalmas a rendszer metaadat-kezelése, és támogatja a legelterjedtebb szabványokat: DC, MARC, MODS, METS, NISO és PREMIS. A Dublin Core mezők száma bővíthető, az egyes gyűjtemények jellemzőihez igazítható. A szakdolgozatok és disszertációk elektronikus változatai esetében például további DC adatelemeket definiáltak a könyvtárosok, hogy a fokozat megnevezését, a tanszék és a konzulens nevét és egyéb olyan információkat is rögzíteni lehessen, amelyeket az NDLTD (Networked Digital Library of Theses and Dissertations) az ETD-MS nevű metaadat-szabványában ajánl. A leíró információk mellett természetesen lehetőség van a hozzáférést szabályozó adatok bevitelére is, a technikai metaadatokat pedig a DigiTool is ki tudja nyerni a fájlokból.

A rendszer egyaránt felkínál egysoros kulcsszavas és összetett keresési lehetőségeket. A digitális objektumokat és a metaadatokat a nagy webes keresőrendszerek is le tudják indexelni. A doku-

mentumok gyűjteményekbe szervezhetőek, melyek azután egy fa-struktúrában böngészhetőek. Az egyes gyűjteményekhez nemcsak egyenként adható hozzá a tételek, hanem automatizáltan is: egy előre definiált keresés találati halmaza is lehet egy részgyűjtemény.

A DigiTool testreszabása

A helyi igényeknek megfelelően a DigiTool egyes elemeit át lehet alakítani. Ezek közül a legfontosabb a felhasználói felület és a metaadatok részletes formátumú megjelenítése. Előbbi egyszerű feladat, de a CSUL-nél szándékosan csak minimális módosításokat végeztek a felületen, elsősorban azért, mert a rendszer időnkénti frissítései felülírhatnak bizonyos fájlokat, amelyeket azután újra módosítani kellene. Más könyvtáraknál is hasonló döntést hozhattak, ezért van az, hogy a DigiTool-t használó intézményeknél a keresőfelület alapvetően ugyanúgy néz ki, csak a színsémák, a menük, a gombok, a fej- és láblécek különböznek. Ami a metaadatok megjelenítését illeti, a rendszer négyféle alternatívát kínál: rövid, táblázatos, teljes és objektum nézet. Utóbbi valamennyi leíró metaadatot megmutatja eredeti formátumban (DC, MARC vagy MODS). A másik három nézetnél szabályozni lehet, hogy milyen adatok jelenjenek meg. A teljes (vagy inkább: részletes) nézet szolgál arra, hogy a felhasználó megítélhesse a dokumentum fontosságát, mielőtt megnyitná azt. Mivel a nagy keresőgépek használóit is ide vezet a DigiTool, ezért különösen fontos volt alaposan végiggondolni, hogy milyen adatmezők jelenjenek meg ezen a weboldalon. (Egy mintapélda a *hdl.handle.net/10217/28672* címen látható.) Ebben a nézetben csak DC metaadatok szerepelhetnek, ezért az egyéb adatformátumokat meg kell feleltetni a DC mezőknek. Bár a DigiTool-ban van ilyen *mapping* a különféle metaadatsémák között, szükség volt itt is némi módosításra a helyi igényekhez és a *Library of Congress* „MARC to Dublin Core Crosswalk” ajánlásához való igazítás érdekében.

A CSUL még a DigiTool beüzemelése előtt igényelt egy intézményi kódot a *handle.net* rendszerben és ezzel a külső szolgáltatással oldotta meg a digitális objektumai stabil azonosítását. Egy ilyen *handle* három részből áll (l. a fenti mintapélda URL címét): egy előtag (pl. *hdl.handle.net*), az intézmény kódja (pl. 10217), valamint egy VIEW típusú objektum azonosítója (pl. 28672), és állandó marad, függetlenül attól, hogy az objektum esetleg idővel már szerverre vagy más rendszerbe kerül át. A *handle.net* azonosítók beépítése a DigiTool-

ba nem volt egyszerű feladat, szükség volt hozzá az Ex Libris szakembereinek segítségére is. A különféle konfigurációs fájlok módosítása mellett azt is meg kellett oldani, hogy a részletes metaadatokban jelenjen meg – kattintható linkként – ez a stabil azonosító, ilyen mezőnévvel: „Bookmarkable URL”. A *handle* nem keletkezik automatikusan a betöltési fázis végén, hanem három lépésben rendelődik hozzá a digitális objektumokhoz. Először generálni kell az azonosítót (ez történhet egyenként a Meditor programmal, vagy csoportosan az objektumok egy halmazára a Management modulal), majd le kell futtatni egy parancsfájlt, ami közlést az azonosítót a külső *handle.net* szerveren és összekapcsolja a DigiTool szerveren levő objektummal. Végül újra le kell aratni a repozitórium metaadatait, hogy a *handle* megjelenjen a felhasználói felületen.

Metaadatok és objektumok betöltése

A DigiTool-ba többféle módon tölthetők be a különböző metaadatok (DC, MARC, MODS, METS) és az egyszerű vagy összetett digitális objektumok. A CSUL jelenleg csak Dublin Core és MARC típusú metaadatokat használ a dokumentumai leírására (amelyek néhány video- és hangfelvételtől eltekintve szinte kizárólag PDF, illetve képek esetén JPEG2000 formátumban vannak). Mivel a DC rugalmas, könnyen előállítható, és mindenféle elektronikus dokumentumfajtahoz jó, ezért ezt alkalmazzák olyankor, amikor új metaadatokat kell készíteni. (A DC fájlok létrehozásához az <oxygen/> nevű XML szerkesztőt használják.) Ha viszont a digitális változatnak van nyomtatott megfelelője, akkor annak a MARC rekordját hasznosítják újra: vagyis csak kiegészítik a szükséges adatokkal, mert így időt és költséget takarítanak meg. Ilyen esetben előbb a katalogizálók kiexportálják a MARC rekordot a Millennium integrált rendszer adatbázisából, majd a nyílt forráskódú MarcEdit szoftverrel kitöltik a fájlnevet és -formátumot, valamint a copyright információt tartalmazó MARC mezőket és almezőket, végül a rekordot MARC XML formátumba mentik és értesítik a repozitóriumot kezelő kollégákat, akik az <oxygen/> programmal validálják az XML fájlt és betöltik azt a DigiTool-ba a digitális objektumot alkotó állományokkal együtt. Amikor nagy tömegben kell metaadatokat előállítani vagy átvenni egy korábbi gyűjteményből, amit a repozitóriumba migrálnak, CSV (comma-separated value) formátumú szövegfájlokat használnak adatcsere céljára, melyekben vesszőkkel elválasztva található az adatsorok. Új metaadatok előállításakor előbb egy Excel tábláza-

tot készítene, amelynek az első sorában a mezőnevek, a többiben pedig az adatrekordok vannak. Az Excel rugalmas szerkesztési funkciói és ismertsége a könyvtárosok között megkönnyíti az adatok bevitelét, melyeket végül CSV formátumba mentenek. A mezőneveket XML fájlok segítségével képezik le (*mapping*) a DigiTool számára, amely így már be tudja konvertálni a CSV állományt az adatbázisába. Már meglévő metaadatok esetében a folyamat annyiban különbözik ettől, hogy a CONTENTdm tartalomkezelő rendszerből előbb tabulátorokkal elválasztott szövegfájlokba mentik a migrálandó metaadatokat, majd ezeket szintén egy leképezési lépéssel alakítják CSV formátumba. Ezzel a módszerrel egy év alatt egy közel tízezer képet, valamint egy hatszáz képet és egyéb dokumentumot tartalmazó gyűjteményt (*University Historic Photograph Collection*, illetve *Water Resources Archive*) tudtak átemelni a repozitóriumba a CONTENTdm-ből.

Összetett objektumok belső szerkezetének definiálására a METS (Metadata Encoding and Transmission Standard) sémát használják. A cikk írásakor 366 ilyen dokumentum volt a repozitóriumban (pl. az egyetem vezetőségi üléseinek jegyzőkönyvei, egy diákújság, illetve egy történeti fotógyűjtemény). A METS objektumok előállíthatók a DigiTool rendszerén belül, vagy azon kívül is, offline módon. Az első esetben is két módszer van: vagy egy „szülő” rekordhoz kapcsolják a betöltött fájlokat (így csak lineárisan navigálható, ún. „fizikai szerkezeti térképek” hozhatók létre), vagy pedig be kell tartani bizonyos fájlnevek konvenciókat (ezzel a módszerrel hierarchikus megjelenésű „logikai szerkezeti térképek” is generálhatók). A leíró metaadatokat a Meditor programmal lehet ilyenkor hozzáadni ezekhez az összetett objektumokhoz, és ugyancsak ezzel lehet szerkeszteni a szerkezeti térképek címkéit. Mivel ez egy időigényes folyamat, csak olyankor használják ezt a módszert, amikor az objektum többféle fájltypusból áll. Ha sok, de azonos formátumú (pl. csak PDF vagy csak JPEG2000) fájl alkotja az objektumot, akkor inkább több lépésben – egy saját fejlesztésű segédprogramot is felhasználva – offline módon készítik el a METS XML-t és azt töltik be a DigiTool-ba.

Integrálás és statisztikák

A DigiTool API-ján keresztül a repozitórium tartalma más felületekről is elérhető. Ilyen például az egyetemtörténeti fotógyűjtemény honlapja (*lib.colostate.edu/archives/uhpc*), valamint a CSU Discovery (*discovery.library.colostate.edu*), ami

egy közös kereső, mellyel a könyvtári katalógusokban, a digitális gyűjteményekben és az egyetemi honlapokon található tartalomban egyszerre lehet keresni (1. ábra). A CSU Discovery a VuFind nevű (PHP és Solr komponensekből álló) nyílt forrású könyvtári OPAC rendszer helyi adaptálása, mely a DigiTool alkalmazásprogramozói csatolóján át gyűjti be a repozitórium metaadatait, majd elmenti őket XML formátumban és betölti a Solr kereső adatbázisába.

A DigiTool adminisztrátori modulja kétfajta jelentést tud készíteni. A Collection Distribution Summary nevű kimutatás a legfelső szintű gyűjtemények méretét listázza, míg a Delivery Usage Statis-

tics a látogatások és a teljesített kérések számát mutatja gyűjteményi és dokumentumszinten egyaránt. Emellett a helyi rendszergazda egy saját statisztikai programot is készített a webszerver naplófájljait kielemezve (kihagyva a keresőrobotokat, illetve a könyvtári dolgozók gépeit), amivel még részletesebb kimutatások és összesítések készíthetők (pl. tanszékenként). Beépítették továbbá a Google Analytics mérőkódját is a repozitóriumba. Ezzel plusz információkat tudnak szerezni, például arról is, hogy honnan érkeznek a felhasználók. 2009. május 26. és június 25. között 5389 látogatója volt a repozitóriumnak, 75,8 százalékuk külső keresőrendszerek (nagyraoszt a Google) találati listájából érkezett.

The screenshot shows the Colorado State University Libraries website. At the top, there are navigation tabs for Home, Collections, Find, Services, About Us, and Help. Below this, there are links for 'For: Undergrads | Grads | Faculty & Staff | Distance | Community | Future | Libraries Staff' and 'My Accounts | Contact Us'. The main content area is divided into several sections: Collections (Catalog (Sage), Regional Catalog, E-Resources & Databases, Digital Collections, Archives), Find (Articles, Books, Journals, Research Guides), and Services. A search bar is prominently displayed with the text 'excavations at the gilligan's island shelters' and a 'Go' button. Below the search bar, there are options to limit the search to 'Library Catalog', 'Digital Repository', or 'CSU Libraries Website'. The search results are displayed in a table with two entries. The first entry is 'Excavations at the Gilligan's Island shelters (5FN1592), Fort Carson Military Reservation (FCMR), Fremont County, Colorado' by Anderson, Cody Mitchell, published in 2008. It is identified as a 'CSU Theses/Dissertations' and is available online. The second entry is 'Excavations at the Gilligan's Island shelters (5FN1592), Fort Carson Military Reservation (FCMR), Fremont County' by Anderson, Cody Mitchell, published in 2008. It is identified as a 'CSU Thesis' and is available in multiple locations. The first entry is highlighted with a red box and labeled 'Returned article'.

1. ábra Keresés a könyvtár honlapjáról a repozitóriumban a CSU Discovery segítségével