

Boros Ildikó

## Digitalizálás – a Digitalizált Törvényhozási Tudástár projekt tapasztalatai

**Az Országgyűlési Könyvtár Digitalizált Törvényhozási Tudástár (DTT) projektje 2010. január 4. és 2012. november 30. között valósult meg, kiemelt projektként, az Elektronikus közgazgatás operatív program (EKOP) keretében, az Európai Unió támogatásával és az Európai Regionális Fejlesztési Alap társfinanszírozásával. A cikk a projekt lelkét jelentő digitalizálás munkafolyamatát mutatja be.**



A projekt fő célja a digitalizálás volt, melynek keretében számos, 1850 és 2010 között megjelent magyar nyelvű könyv, időszaki kiadvány, jogforrás, jogi, történettudományi, országgyűlési és politikai szakirodalom digitalizálása történt meg. A projekt keretében létrejött új portál célja, hogy tartalomszolgáltatásával hatékonyan segítse a jogalkotási munkát, a jogalkotás háttérében folyó szakmai tevékenységet, támogassa a felsőoktatásban és a tudományos műhelyekben folyó oktató- és kutatómunkát, a könyvtárak jogi, politológiai és történettudományi tájékoztató tevékenységét és a megszerzett tapasztalatokkal a könyvtárak jövőbeni digitalizálási tevékenységét.

### A projekt szakaszai

1. Megvalósíthatósági tanulmány benyújtása (2008)
2. Pályázati, tervezési szakasz (2009. június – 2010. január)
3. Szerződés-kötés (2010. április 13.)
4. Kommunikációs, előkészítő szakasz (2010. január – június)
5. Digitalizálási közbeszerzési eljárás (2010. május – 2011. január)
6. Digitalizálás - végrehajtási szakasz (2011. január 21. – 2011. december)
7. Portál közbeszerzési eljárás (2011. május – 2011. december)

8. Portál, tartalomszolgáltatás - végrehajtási szakasz (2011. május – 2012. szeptember)
9. Adatbázis-publikálás, tárgyszavazás (2011. június – 2012. október)
10. Kommunikációs, záró szakasz (2012. szeptember – 2013. június)
11. Projektzárás (2012. november 30.)
12. Fenntartási időszak (2012. november 30. – 2017. november 30.)

### A digitalizálás előkészítése

#### Az informatikai háttér megteremtése

A projekt egyik meghatározóan fontos eleme az informatikai környezet kialakítása volt: a hardver- és szoftverszükségletek, illetve az archiválás tervezése. Az így megvalósult bázis a könyvtárban folyó hosszú távú digitalizálás alapjait kívánta megteremteni. A digitalizált állomány tárolása és szolgáltatása érdekében az *Országgyűlés Hivatala* közbeszerzési eljárás keretében, saját költségvetésből, a projekt *hardver* háttérének támogatása céljából korszerűsítette szerverparkját, és megvásárolta a létrejövő digitális gyűjtemény kezelését biztosító *DigiTool* szoftvert. A DigiTool ún. digitális gyűjteménykezelő szoftver, amely minden fajta (szöveges, audio és video) digitális médiát kezel. A digitális állomány a DigiTool-ban kiegészíthető a digitális dokumentumokhoz tartozó metaadatokkal.

Ezt követően kerültek meghatározásra a digitalizálás *általános műszaki paramétereit*, amelyek a következők:

- *Képfarmátum*: TIFF (Tagged Image File Format). A TIFF a digitális képek szabványos tömörítetlen formátuma. Mivel ebben a formátumban nincs tömörítés, a kép minősége megfelelő.

- *Felbontás:* 400 dpi
- *Színmélység:* 8 bit
- *Oldalkép:* Egy kép csak egy dokumentumoldalt, a teljes oldalt tartalmazza.
- *Margók mérete:* A margók mérete nem kisebb 50 pixelnél és nem nagyobb 200 pixelnél.
- *Oldalferdeség:* Az oldalferdeség nem haladhatja meg a 0,5 fokot.
- *Mérethűség:* Az oldalak mérethűen, az eredeti kötetnek megfelelően lettek elmentve (1:1 arányban). Az oldalméretek beállítása kötetenként azonos, a dokumentumok oldaltájolásán nem változtattunk.
- *Bélyegkép:* Minden digitalizált dokumentum címlapjáról bélyegkép készült.
- *Képjavitó eljárás:* Minden oldal arányos körbevágása megtörtént.
- *Karakterfelismerési minőség:* (OCR = Optical Character Recognition). A főszöveg esetében: 95%-os minőség, azaz 100 karakter maximum 5 hibát tartalmazhat; speciális oldalak esetében: 995 ezrelékes pontosságú, azaz 1000 karakterből 5 lehet hibás; problémás oldalak esetében: 90%-os minőség, azaz 100 karakter maximum 10 hibát tartalmazhat.
- *Karakterkészlet/kódolás:* UNICODE/UTF-8
- *OCR formátum:* Az elkészült oldalak kereshető, kétrétegű PDF formátumban kerültek feltöltésre a DigiTool szoftverbe.
- *Metaadat:* Minden oldalt az oldal tartalma szerint egyedi azonosító jelzettel láttunk el: szöveg, kép, tárgymutató, oldalszám, többhasábos szöveget tartalmaz-e, az oldalon szereplő szöveg nyelve, a kiadvány lapszáma, évszám, egyéb azonosítók. (A metaadatok hozzárendelése a művekhez a keresés elősegítése érdekében szükséges.)
- *Feldolgozott állomány átadása:* 2 TB kapacitású külső merevlemezekben.
- *Kötetek azonosítása:* Az egyes kötetek azonosítása vonalkód alapján történt.

Következő lépésként megtörtént a *digitális dokumentumok fizikai és logikai megjelenítésének kialakítása* – az alábbiak szerint: könyvek (lehetnek egy- vagy többkötetesek, tartalomjegyzékkel vagy a nélkül); döntvénytárak (lehetnek egy- vagy többkötetesek); időszaki kiadványok (cikkek), közlönyök (füzetek). Alapja: METS (Metadata Encoding and Transmission Standard = metaadat-kódolási és átviteli szabvány); adatscsoportok: fejlécszöveg, leíró metaadat, adminisztratív metaadat, objektumcsoport, strukturális térkép.

Ezt követően kialakításra került a *tömeges betöltés* munkafolyamata. Lépései: adatbázis-építés a szkennelt fájlok adataiból (fájlnév, útvonal), dokumentumcsoportonként (közlönyök, döntvények, könyvek, folyóiratok). Összefűzött PDF állomány (cikkek, füzetek, kötetek stb.) készítés az adatbázis adatainak segítségével. Ehhez segédlet készült: nevezéktan dokumentumcsoportonként, dokumentáció a fizikai és logikai megjelenítésről a DigiToolban és további segédtablák az alábbiak szerint (az adatok a könyvtár adatbázisaiból előállíthatók):

- Közlönyök: rendszerszám, jelzet, közlönycím, év, copyright;
- Döntvények: rendszerszám, vonalkód, döntvénycím, év, copyright;
- Könyvek: rendszerszám, vonalkód, könyvcím, év, copyright;
- Folyóiratcikkek: rendszerszám, jelzet, folyóiratcím, év, füzet, oldal, cikkcím, copyright.

Dokumentumcsoportonként létrehozott sablonok alapján METS/XML készült, a könyvtári adatbázisokból nyert adatokból.

A matematikai statisztika módszereivel előzetesen megtervezték a digitalizált állomány *minőségellenőrzését*. A kiválasztott eljárás az érvényes MSZ 548-77 és ISO 2859-1 szabványok I-es fokozatú ellenőrzésnek elvégzésén alapult az alapsokaság mérete alapján, mintavételes eljárással.

### A művek kiválasztásának alapelvei

A projekt alapvetően fontos eleme volt a művek kiválasztása digitalizálásra. A digitalizálást megelőzően *általános alapelveket* határoztunk meg, amelyek tudományterület, dokumentumtípus, nyelv, idő stb. vonatkozásában határozták meg a művek csoportjait. A műveket a jogtudomány, a történettudomány és a politikatudomány területéről, vagyis az Országgyűlési Könyvtár főgyűjtőköreinek alapvető szakirodalmából válogattuk ki. A kiválasztott művek mind magyar nyelvűek. Tekintettel arra, hogy fontos cél volt a digitalizált művekhez való széles körű, internetes hozzáférés biztosítása, arra törekedtünk, hogy a listában többségben legyenek a szerzői jogi védelem alá már nem tartozó művek. Ennek megfelelően főként 1850 és 1945 között megjelenő műveket választottunk ki; az 1945 utáni szakirodalomból kisebb mennyiséget. Az egyes tudományterületekhez tartozó listákat szakterületi munkacsoportok állították össze.

Az előkészítési feladatok közé tartozott az ún. *tartaléklisták összeállítása*. Ebben a munkafázisban kb. 7500 kötet (2 800 000 oldalnyi terjedelemben) előkészítése történt meg.

### A kötetek előkészítése

A könyvtár előzetes állapotfelmérésben rögzítette a kötetek terjedelmi, szerkezeti és átadás kori állapotának legfontosabb jellemzőit (ún. állapotfelmérő kísérőlapon). A könyvekről egyedi, a folyóiratokról, közlönyökről és a döntvényekről csoportos állapotleírást készült. A kísérőlap elkészítését kitöltési útmutató segítette. Az eredmények rögzítése ún. állapotfelmérő adatbázisban történt.

A *könyvek egyedi állapotfelmérését* az ún. Állapotfelmérő munkacsoportot végezte el. A munkacsoport feladata volt a digitalizálásra kerülő dokumentumok könyvészeti, fizikai és könyvkötészeti szempontok szerinti értékelése. Az értékelés célja az volt, hogy

- minden egyes kötet vonatkozásában eldöntsék, milyen digitalizálási technológiát lehet és szabad alkalmazni;
- regisztrálják, hogy az adott kötet milyen szkennelési, szövegfelismeretési problémákat jelenthet;
- amennyiben lapra vágásra kerül a kötet, a helyreállítás érdekében pontos képünk legyen az eredeti könyv szerkezetéről, kötéséről;
- fel tudjuk mérni az előkészítési és helyreállítási munkákat;
- ki lehessen válogatni a problémás és problémamentes köteteket annak érdekében, hogy viszonylag egységes bánásmódot igénylő csomagokat lehessen összeállítani a digitalizáláshoz.

Az értékelés az alábbiakra terjedt ki:

- a *mű*, majd a *példány* kiválasztása;
- a *kötés vizsgálata*: nyithatóság, belső margó, borító, kötés;
- a *könyvtest vizsgálata*: terjedelem, méret, mutatók, mellékletek stb.;
- *hiányok*: összegyűjtése, pótlása;
- a *papír vizsgálata*: elszíneződés, szálirány, töredezettség;
- a *szövegtűkőr vizsgálata*: átnyomódások, olvasói bejegyzések, aláhúzások, eltérő szövegírány, táblázatok, foltok, lyukak, szakadások.

Az *időszaki kiadványok* esetében *csoportos állapotfelmérés* történt, az adatgyűjtés és az értékelés az alábbiakra terjedt ki:

- bibliográfiai leírás, előzmény, folytatás, címváltozás;
- évfolyam; megjelenési év;
- átadott kötetek (fizikai egységek) száma /db.;
- állapotleírás;
- éves tartalomjegyzék, helye;
- füzetszám;
- oldalszámozás típusa (folyamatos vagy újrakezdődő);
- füzetborító helye;
- füzetszámok tartalomjegyzéke, helye, oldalszáma;
- melléklet címe, száma, helye;
- eltérő méretű oldalak helye, oldalszáma;
- hiányzó oldalak, hiányzó számok helye, oldalszáma;
- téves oldalszám, téves kötészet sorrend helye, oldalszáma;
- OCR problémák; utómunka;
- oldalszámok: melléklet és a teljes évfolyam oldalszám.

Az Állapotfelmérő kísérőlap egy másolati példánya végig kísérte a köteteket.

### A közbeszerzési eljárás

A feladat tervezése során dönteni kellett arról, hogy a digitalizálást külső vállalkozóval vagy saját munkával célszerűbb-e elvégezni. A korábbi tapasztalatok, valamint az előnyök és a hátrányok mérlegelése után a könyvtár a külső vállalkozóval történő munkavégzés mellett döntött: a munka mennyisége, a rendelkezésre álló rövid idő és a piacon megjelenő robottechnológia újdonsága egyaránt egy professzionális külső vállalkozás bevonása mellett szólt. A könyvtár munkatársai az előkészítési feladatokat végezték el.

Első lépésként kidolgozásra került a digitalizálási közbeszerzés műszaki-szakmai specifikáció, amelyben tartalmazta az elbírálási szempontokat, a műszaki-szakmai, és a gazdasági-pénzügyi alkalmassági feltételeket. Meghatározásra kerültek a szkenneléssel, a karakter-felismeretéssel, a fájlnevezésekkel, a digitalizált állomány betöltésével, a kötetek átadásával, visszavételével és a kötetek állományvédelmével kapcsolatos elvárások, továbbá bemutatásra kerültek a digitalizálásra kiválasztott kötetek jellemzői, kiemelve mindazon információkat, amelyek meghatározóan fontosak a szkennelés és a karakter-felismeretés folyamatában.

2010 júniusában jelent meg az első ajánlattételi felhívás, amely alapinformációkat, műszaki-szakmai specifikációt és jogi követelményeket tartalmazott. Minden pályázónak számos jogi követelménynek kellett megfelelni: felolvasólap készítése; nyilatkozat a kizáró okokról; a köztartozásmentes adózásról; az előző három év árbevételeiről; nyilatkozat az előző három évben végzett könyvtári vagy levéltári dokumentumok digitalizálására, karakterfelismeretetésére, adatbázisba szervezésére vonatkozó szolgáltatás teljesítéséről; a szerződésben részt vevő szakemberekről; a rendelkezésre álló erőforrásokról (a digitalizálásra alkalmas gépekről); a biztosítékokról; a részteljesítésekről; a teljesítésbe bevont alvállalkozókról, szakemberekről. Az érdeklődőknek egy ún. helyszíni bejárás keretében bemutattunk egy válogatást a digitalizálásra váró kötetekből. Ezt követően az eljárásban részt vevő cégek benyújtották ajánlataikat, amelyeket az ajánlattételi felhívásban meghatározott szempontok alapján értékelve és megtörtént a szerződéskötés a nyertessel. A közbeszerzési eljárás az előzetesen tervezettnél sokkal több időt vett igénybe, mivel az első eljárás eredménytelenül zárult. A második közbeszerzési eljárás nyertese a *Bravogroup Rendszerház Kft.* és a *DocuScan Dokumentumkezelő Kft.* konzorciuma lett.

A digitalizálást a DocuScan Dokumentumkezelő Kft. végezte. A cég 15 éve foglalkozik dokumentumok kezelésével, digitalizálásával, jelentős tapasztalattal rendelkezett a digitalizálás területén, ám ebben a projektben egy teljesen új technikát kellett alkalmaznia. A korábbi, főként nagyteljesítményű lapbehúzó szkennerekre épülő, különálló lapokat kezelő feldolgozási munkafolyamatról át kellett állni a robotszkenner segítségével végzett könyvdigitalizálásra. Speciális feldolgozási eljárást igényelt a digitalizált állományok adatbázisba szervezése és portálon való szolgáltatása is.

A digitalizálási közbeszerzésből közvetlenül adódó feladatok elvégzése a könyvtár létrehozott ún. Digitalizálási munkacsoportot.

### **A kötetek átadásának és átvételének előkészítése**

A Digitalizálási munkacsoport egyik feladata volt a kötetek folyamatos biztosítása a digitalizáláshoz, a három műszakos, tömeges feldolgozáshoz. Kidolgozásra került a kötetek átadásának és visszavételének munkafolyamata és dokumentálási rendje. A könyvtári kötetek átvétele tételes átadás-átvételi

jegyzőkönyvvel történt, csomagokban. A könyvtár az állomány védelme és megóvása érdekében külön állományvédelmi tájékoztatót készített, amely tartalmazta a kötetek tárolásával, szállításával, lapozásával kapcsolatos tudnivalókat.

### **A metaadatok előkészítése**

A megfelelő példányok kiválasztása céljából szükséges volt a digitalizálásra javasolt művek könyvszerzeti, példány- és kiadási adatainak behasonlítása is. Elengedhetetlen volt a könyvek és az időszaki kiadványok adatbázis rekordjainak ellenőrzése, javítása, előkészítése konverzióra, a szerzői jogi státusz és a gyűjteménybe szervezési kódok bevitelle az adatbázisba. Megtörtént a metaadatok előkészítése. A digitálisan elérhető dokumentumok adataihoz használt szabványok a következők voltak: Dublin Core ajánlások (DCMI). Metaadat leíró szabvány. MSZ ISO 15836. ESE – Europeana Semantic Elements. Version 3.4.1.

Elkészült a metarekordok bibliográfiai adatelem specifikációja is. A digitalizálandó folyóiratok cikkeinek egy részéről rendelkezett a könyvtár rekordokkal, a hiányzókat a Vállalkozónak kellett elkészítenie. E munkafolyamat általános és az adatok leírásának specifikus szabályait az Adatbázis munkacsoport dolgozta ki.

Az előkészítő munkafázisban készült a digitalizált könyvek tartalomjegyzékeinek rögzítése; külön munkacsoportok foglalkoztak az új portál tartalom-szolgáltatási alapelveinek kialakításával és a dokumentumokhoz való szerzői jogi hozzáféréssel.

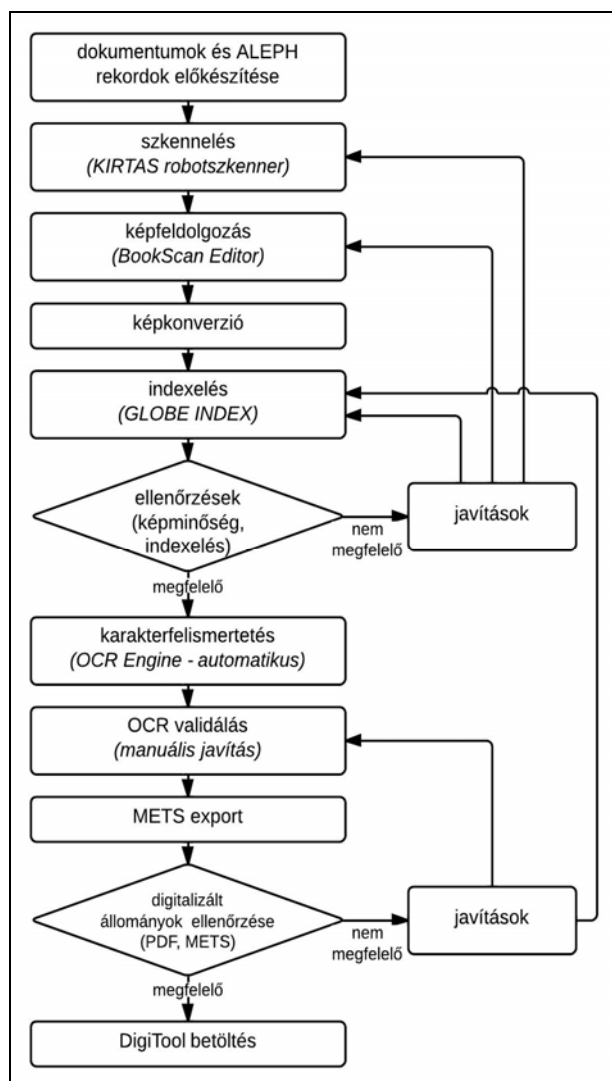
## **A digitalizálás**

### **A digitalizálás közös irányítása**

Az együttműködés szabályaira vonatkozóan a nyertes vállalkozó és a könyvtár egy ún. *Projekt Alapító Dokumentumot* (PAD) készített, amely tartalmazta a közös munkára vonatkozó eljárásokat, a működési szabályokat, a digitalizálás ütemtervét, a szervezeti renchez kapcsolódó felelősségi és hatásköröket, a vezetéssel kapcsolatos eljárásrendet, továbbá a közös munka dokumentálásának rendjét (1. ábra).

A digitalizálási feladatok és az egyes szakmai csapatok munkájának gyors nyomon követésére, valamint az operatív munkát befolyásoló döntések meghozatala céljából a könyvtár és a vállalkozó

heti rendszerességgel ún. Projektvezetői megbeszélést tartott. A munka során készített elektronikus és papír dokumentumok tárolására és visszakereshetőségének biztosítására, a munkadokumentumok különböző verzióinak kezelésére közös dokumentumkezelő rendszer, a *Therefore* program került bevezetésre, közös, egységes sablonok használatával (emlékeztető, átadás-átvételi igazolás, teljesítési igazolás, státuszriport, változáskezelő-sablon).



1. ábra A digitalizálás folyamata

## A digitalizálás helyszíne

Tekintettel a digitalizálásra kiválasztott művek értékére, egyediségére, sok esetben pótolhatatlanságára, valamint az állomány védelmére, a digitalizálás *helyben* történt: az Országház épületében, a

könyvtár területén erre a célra elkülönített helyiségekben, a könyvtárról leválasztott és a Vállalkozónak átadott területen került sor a digitalizálásra és a kapcsolódó vállalkozói háttérfeladatok elvégzésére. A tesztüzemet követően a digitalizálást végző cég három műszakos munkarendben szervezte és végezte feladatait.

## Tesztüzem

2011. február 2-ától március 3-áig tartott az előzetes tesztelés időszaka. Ez idő alatt minden dokumentumcsoport tekintetében részletes előzetes konzultáció zajlott a Könyvtár és a Vállalkozó képviselői között. A tesztcsomag előkészítéshez külön szempontrendszert készült, mennyiségi, minőségi, szkennelési és fájlnevezési szempontokat is figyelembe véve. A tesztidőszak alatt minden dokumentumcsoportból megtörtént a kiválasztott dokumentumok szkennelése, optikai karakterfelismertetése, indexelése és feltöltése a DigiTool rendszerbe. Ennek célja a munkafolyamat optimalizálása, a vállalkozóval való egyeztetés, a rekordok összekapcsolásának és a betöltés módjának a meghatározása, a használati jogosultságok kijelölése volt.

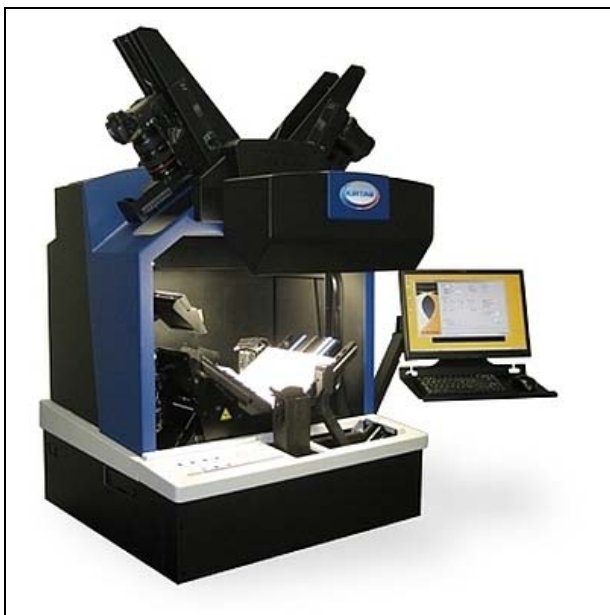
## Szkennelés

A szkennelés nagy része *Kirtas KABIS III.* típusú *robotszkennel*rel történt (2. ábra). A robotszkennel jobb és bal oldalán két, egyenként 21.1 megapixel felbontású Canon típusú kamera található, amelyek a kötetek bal és jobb oldaláról készítik felvételeket a kívánt formátumban. A szkennel maximum névleges kapacitása 3000 oldal/óra. A maximum kötetméret: 28x35.6 cm. A szkennelés sebességét alapvetően a kötetek állapota határozta meg, ennek megfelelően lehetett kiválasztani az alkalmazott szkennelési eljárást.

A legjobb minőségű kötetek *automata üzemmódban* szkennelhetők, amely esetben az operátorok a kötet gépbe helyezését és alapbeállítások elvégzését követően „csak” felügyelték a gép működését. A lapozást és a fotózást a gép automatikusan végezte. A gép teljesítménye – tapasztalati értékek alapján – 25 000 oldal/nap (24 óra).

A rosszabb állapotú, sérült vagy régi kötetek esetében a folyamatot manuális rásegítéssel kellett elvégezni, előfordult a szkennelés során, hogy szükséges volt minden egyes oldal igazítása. *Manuális üzemmódban* a szkennelési sebesség –

tapasztalati értékek alapján – 10-13 000 oldal/nap (24 óra).



2. ábra Kirtas KABIS III. típusú robotszkenner

A nagyméretű köteteket, a kihajtható mellékleteket *síkágyas szkennerrel* készítették el.

### Képfeldolgozás

A munkafolyamat következő lépése a *képek feldolgozása, vágása és javítása* volt. Ebben a folyamatban ún. *Book Scan Editor (BSE)* szoftver segítségével végezték el a képfeldolgozáshoz szükséges beállításokat például file-ok elnevezése, egy ferdén elhelyezkedő oldal kiegyenesítése, a szkenner lefogóinak kivágása, fény- és kontrasztarányok beállítása, szövegtűkör körüli vágókeret beállítása stb. Ezeket a beállításokat a program egy XML típusú file-ba mentette. Ennek a file-nak a segítségével tudta az automata vágó legyártani a megfelelő TIFF-formátumú képeket. Ez a munkafolyamat lassú; tapasztalati értékek szerint átlagosan kb. 8 kötet készült el óránként.

Az automata vágás az ún. Super Batch programban történt, ahová be kellett tölteni a kötethez tartozó nyers képeket és automatizált módon megvágni. A vágás beállító által megadott és XML fájlba mentett paraméterein alapján a program TIFF-formátumú fájlokat generál.

Az automata vágás által létrehozott képek javítása során lehetőség nyílt minden oldal ellenőrzésére.

## Boros I.: Digitalizálás – a Digitalizált Törvényhozási ...

A művelet során a vágás-beállító által beállított paraméterein lehetett finomítani, változtatni (pl. ferdeség, vágókeret stb.).

A könnyebb kezelhetőség érdekében a TIFF tömörítetlen képeket JPEG tömörített formátumú fájlokká konvertáltuk, így 5-30 MB méretű TIFF-kép helyett 100-600 KB-ra leegyszerűsített, ún. *nézőképekkel* lehetett dolgozni.

Az ellenőrzés folyamán került kijelölésre a kötetek első 10 oldalából az a felvétel, amely a *bélyegkép* lett.

### Metaadat-készítés

Ebben a munkafolyamatban zajlott a GLOBE-Index nevű program segítségével a metaadatok hozzárendelése minden kötet minden oldalához. Ilyen információk, például az oldal tartalma szerint: szöveg, kép, tárgymutató, oldalszám, több hasábos szöveget tartalmaz-e az oldal, a szöveg nyelve, a kiadvány lapszáma, évszám, egyéb azonosítók.

### Karakterfelismertetés

A szkennelt képek esetében szövegkonverziós eljárás, optikai karakterfelismertetés (*OCR = Optical Character Recognition*) biztosítja a kereshetőséget. Ez a technológia, bizonyos hibagyakoriság mellett, képes automatikusan felismerni a nyomtatott oldalon látható karaktereket (betűket). Azoknak az oldalaknak az esetében, ahol a könyvtár magas minőségi elvárást határozott meg, manuális javítás vált szükségessé. A karakterfelismertetés a munkafolyamat szintjén azt jelentette, hogy a képfeldolgozás után az adatok továbbkerültek az OCR adatbázisba, s egy teljesen automata folyamatként az ún. *OCR Engine* létrehozta a már végleges formátumú kétrétegű PDF állományt.

### Tömeges feltöltés

Az elkészült oldalakat kereshető, kétrétegű PDF formátumban töltötték fel a DigiTool szoftverbe. A folyamat végén a program rendszerezte és összemásolta a PDF-eket egy ún. DigiTool-állományba, amely már a végleges betöltés tárgya. A METS export folyamán olyan betöltő csomag jött létre, amely tartalmazta a PDF-et, a kötet belső struktúráját, és lehetővé tette, hogy a katalógusból a bibliográfiai adatokat az objektumhoz kapcsoljuk.

## Ellenőrzés

Az ellenőrzési munka első fázisában a vállalkozó teljesítési késedelme miatt csak az első átadások nagy tömegű ellenőrzését követően váltak láthatóvá a típushibák és azok nagyságrendje. Ennek eredményeképpen a hibák javítása nagyobb mennyiségű oldalt érintett.

Az ellenőrzés első lépése a TIFF-képek automatikus ellenőrzése volt. A szkennelés folyamán előállított nyersanyag, TIFF-fájl, a későbbi eljárások alapja, ezért ezen a ponton a könyvtár minőségellenőrzést végzett.

Az *automatikus ellenőrzés* a képek technikai paramétereinek (formátum, felbontás, színmélység, minimum-maximum méret) és mennyiségének vizsgálatára terjedt ki. A talált hibák minősége és mennyisége miatt a könyvtár manuális ellenőrzést is végzett.

A *manuális ellenőrzés* folyamán a szkennelés általános minőségellenőrzése zajlott (élesség, ferdeség, hullámosság, körbevágás, oldalsorrend, oldalhiány, dupla oldalak, margók szélessége, mérethűség szempontjából), illetve a folyóiratok és közlönyök esetében annak vizsgálata, hogy megfelelő-e a kötetjelzés, a megjelenési év, a füzetszám, a mellékletek, a füzetborítók, a tartalomjegyzék, a mutatók elnevezése, azonosítása, pontos-e az egyes adatok felvétele.

Az ellenőrzés technikai és adminisztratív feltételeinek kialakítása érdekében a számítógépek adottságaihoz igazodva megtörtént a képfájlokat kicsinyítése/tömörítése. A képanyagok (TIFF-képek) ellenőrzése 5 vagy 10 ezer oldalt tartalmazó csomagokban történt. Az ellenőrzéshez egységes szempontrendszert használtunk, külön a könyvekre és külön az időszaki kiadványokra. Tömeges hibák észlelése esetén rendkívüli vizsgálat elvégzésére volt szükség, majd javaslatot kellett kidolgozni a kockázatok további elkerülésére.

Azok a kötetek, amelyek megfeleltek a minőségi követelményeknek, „OK” minősítést kaptak, majd átadtuk őket a következő munkafázisra. Ha a hibák mennyisége a kötet oldalszámához viszonyítva meghaladta a 10%-ot, akkor a teljes kötetet újra feldolgozták. Ha a hibák mennyisége nem haladta

meg a 10%-ot, abban az esetben csak a hibás (egy-egy) oldalak javítása történt meg.

## A digitalizált állomány tárolása

Az archiválás IBM Tape Library kazettás egység passzív háttérkapacitású DAT-kazettára történő archiválás és mentési célú tárolás HP Data Protector rendszerrel. A digitalizált dokumentumok eredeti képei (TIFF-fájlok) hozzávetőleg 15 TB tárhelyet foglalnak el. Az eredeti képfájlok alapján készült kétrétegű PDF-fájlok a központi háttértárolón jelenleg 3 TB-ot foglalnak el.

## Összegzés

A projektben digitalizált oldalak száma 2 millió oldal, a kötetek száma összesen 5272 kötet. A digitalizált dokumentumokhoz a hatályos szerzői jogi szabályozásnak megfelelően lehet hozzáférni. A művek jelentős része (mintegy 40%) azonban szerzői jogi védelem alatt áll, így csak az Országgyűlési Könyvtárban elhelyezett számítógépeken jeleníthető meg, tudományos kutatás vagy egyéni tanulás céljából. A szerzői jogvédelem alá nem tartozó művek a nyilvánosság számára szabadon hozzáférhetők az interneten. Ez esetben a dokumentumok egésze vagy része letölthető, menthető és nyomtatható. A DTT-portál informatikai szempontból akadálymentes, a gyengén látók számára is jól használható.

A DTT projekt keretében digitalizált dokumentumokat digitális gyűjteményekbe rendezetten, a DigiTool nevű gyűjteménykezelő szoftveren keresztül szolgáltatjuk. A digitalizált köteteket a hierarchikus struktúrában rendezett gyűjteményekben lehet áttekinteni, böngészni. A gyűjteményeken belül további algyűjtemények segítik a tájékozódást.

A DTT nagyobb gyűjteményei a következők: jogforrások, könyvek, folyóiratok.

- *Jogforrások: Közlönyök:* 11 féle közlöny, 1129 kötetben, 20 924 füzetben; *Döntvények:* 39 féle döntvénytár, 605 kötetben.
- *Könyvek:* 864 mű, 2248 kötetben.
- *Folyóiratok:* 73 féle folyóirat, 1674 kötetben, 20 800 füzetben.

A digitalizált gyűjtemények, hozzájuk kapcsolódó tudástárral az ún. DTT-portálon (3. ábra) átérhetőek el.

3.ábra A DTT projekt honlapja: <http://dtg.ogyk.hu/hu/>

Beérkezett: 2013. V. 20-án.

**Boros Ildikó**  
 az Országgyűlési Könyvtár  
 Gyűjteményszervezési osztályának  
 vezetője.  
 E-mail: [boros.ildiko@ogyk.hu](mailto:boros.ildiko@ogyk.hu)

## Ingyenesen letölthető szerzői jogi kézikönyv

A ProArt Szövetség a Szerzői Jogokért honlapjáról ingyenesen letölthető a *Kézikönyv a szerzői jog érvényesítéséhez* című kiadvány.

A szerzői joggal foglalkozó szakembereknek szóló hasznos kiadványhoz juthat, aki honlapunkról letölti a *Kézikönyv a szerzői jog érvényesítéséhez* című kötetet. A mű a jogkereső jogosultak, felhasználók, hivatásos és alkalmi jogalkalmazók számára készült tudományos gondossággal, pontossággal.

Az első fejezet a szerzői jog mindenkor hatályos anyagi jogi szabályozásába enged betekintést, ismertetve a bírói és szakértői gyakorlatot, valamint az Európai Bíróság döntéseit, majd a polgári jogi jogérvényesítés eszköztárát mutatja be, rávilágítva a legjobb gyakorlatokra és jogérvényesítési problémákra is. Ezt követi a büntetőjogi jogérvényesítési fejezet, elsősorban a Btk. rendelkezéseire koncentrálván. A kötetet a vámjogi fejezet zárja, amely terület ugyan a szerzői jogi jogérvényesítésben egyre kisebb szerepet játszik, a jogérvényesítési kötet azonban semmiképp sem lehetne teljes nélküle.

Jó olvasást kívánunk!

ProArt Szövetség a Szerzői Jogokért

<http://konyvtar.bpugyvedikamara.hu/2012/11/ingyenesen-letoltheto-szerzoi-jogi-kezikonyv>