

A Magyar Távirati Iroda Hírchívumának digitalizálása

A megfelelő és hiteles írott források felkutatása még napjainkban is bonyolult és időigényes feladat, és az iratanyagok hozzáférhetetlensége miatt gyakran ütközik nehézségbe. Mindez nem csupán a múlt században keletkezett, feldolgozatlan tartalmak esetében igaz. A nem megfelelő kezelés következtében például számos, az elmúlt évtizedekben született írásos dokumentum is elérhetetlenné vált, így kulturális szempontból nélkülözhetetlen művek merültek feledésbe. A digitalizálás azonban megoldást nyújt az értékes kulturális tartalmak archiválásában és minden állampolgár számára, ingyenesen, hozzáférhetővé teszi az információkat.

Közel 20 éve a hazai tartalomszolgáltatói piacon

Az 1989-ben alakult Arcanum Adatbázis Kiadó és Fejlesztő Kft. a hazai CD-ROM piac egyik úttörőjeként adatbázisok építésével, kiadásával foglalkozik. Az Arcanum küldetésének tekinti, hogy elérhetővé tegyen olyan kultúrtörténeti értékű, könyv alakban már nem vagy csak nehezen hozzáférhető műveket a jövő olvasói számára, mint például a Magyar Távirati Iroda Hírchívuma. Arra törekszik, hogy technológiai tudását a magyar és egyetemes kultúra hasznára fordítva, olyan offline és online kiadványok sorát jelentesse meg, amelyek a jövő digitális könyvtárainak alapművei lehetnek.

Az 1881 óta működő magyar nemzeti hírügynökség, a Magyar Távirati Iroda legfontosabb feladatának a gyors, pontos és megbízható hírszolgáltatást tekinti. Az összes megyére kiterjedő tudósítói hálózatának köszönhetően az MTI az ország egész területéről gyűjti az információkat. Mivel egy-egy hír gyakran csak a későbbiekben válik fontossá, a hírügynökség kiemelt hangsúlyt fektet minden információ archiválására, hogy azok bárki által könnyedén visszakereshetők legyenek.

A Magyar Távirati Iroda szöveges dokumentumainak digitális online közzétételi feladatai az 1920 és 1949 közötti időszakra, valamint az 1956-ban keletkezett dokumentumokra terjedtek ki.

A Hírchívum digitalizálásának kihívásai

A minőségből, jellegből és állagból adódó problémák több szempontból is indokolták a dokumentumok digitalizálását. Egyrészt erősen korlátozott volt a károsodott állapota miatt védett anyagnak minősített Hírchívum hozzáférhetősége, elveszítve ezzel lényegi funkcióját. Másrészt a hatalmas, írógéppel készült és nyomatos technikával sokszorosított gyűjteményt nem lehetett állagmegóvó eljárás alá vonni, mivel az idők során visszafordíthatatlanná vált károsodási és savasodási folyamatok következtében megsemmisülhetett volna. Jelenlegi állapotában azonban még alkalmas volt a digitális feldolgozásra.

A feladatot tovább nehezítette, hogy mindössze két hónap állt rendelkezésre a lapokra szedett, egyoldalas, rendkívül rossz minőségű, sárgult, egyenetlen és nehezen olvasható szöveg digitalizálására, amelyet a levéltárban, a nyitvatartási időben lehetett csupán elvégezni.

A kihívást tehát egy több százezer oldalnyi, 1920 és 1949 között, valamint 1956-ban keletkezett, mára azonban jelentősen megrongálódott gyűjtemény gondos és kíméletes munkával történő digitalizálása jelentette. Így a sikeres kivitelezésnek köszönhetően új távlatok nyílhattak meg a kulturális tartalmak elérhetőségében.

Kiemelkedő minőségű eszközpark

A körülmények sokszoros nehézséget jelentettek a projekt végrehajtása során, ahol a határidő rövidege miatt a munkafolyamatokat párhuzamosan és nagyon feszes tempóban kellett elvégezni. Ehhez a munkához olyan nagyteljesítményű szkennerekre volt szükség, amelyek napi több tízezer oldalas kapacitásuk mellett nagyon kíméletesen bánnak a digitalizálandó anyaggal, és megbirkóznak a jelentősen károsodott papírokkal is.

A Hírárchívum digitális közzététele kizárólag speciális eszközpark segítségével valósulhatott meg. A Canon professzionális digitalizáló eszközpalettája hatékony segítséget kínált a projektet kivitelező Arcanum Adatbázis Kiadó számára, hiszen a vállalat 3-4 eszközből felállított egy komplett digitalizáló stúdiót, amely szinte minden feladatra hatékony megoldást nyújtott. Az innovatív eszközpark biztosítja, hogy az Arcanum a szöveges adatbázisokkal párhuzamosan képi és térkép-adatbázisokat is elérhetővé tegyen elektronikus formában, s jelenlegi kapacitásával évi 2-3 millió szövegoldal és 300-400 ezer képi anyag digitalizálására, valamint publikálására képes.

A hatalmas volumenű digitalizálás során keletkezett por miatt a gépek folyamatos tisztítása mellett minden egyes lapra külön figyelmet kellett fordítani. A szkennerek – amelyeket a rendkívül vékony és porózus anyag megóvása miatt csak a legkíméletesebb üzemmódban lehetett használni – lehetővé tették a Magyar Országos Levéltár által rendelkezésre bocsátott sokszázezer oldalnyi Hírárchívum szinte sérülésmentes digitalizálását.

800.000 oldalnyi dokumentum digitalizálása 40 munkanap alatt

A sikeresen megvalósult projekt során több mint 800.000 oldal digitalizálására került sor – a gyors munkatempónak, jó szervezésnek és hatékony szkennereknek köszönhetően – mindössze 40 nap alatt. Ezek a dokumentumok kereshető formában kivétel nélkül publikálásra is kerültek. Az Arcanum szakemberei napi nyolcórás munkaidőben, két szkennel segítségével digitalizálták a Hírárchívumot, amelynek eredményeképpen létrejött digitális állomány várákoszáson felüli minőségű lett. Az adatbázis – amely a mai helyesírási szabályok szerinti egységes keresést követi – felkínálja az összes felismert szó listáját, az esetleges hibák könnyedén észrevehetőek, így a keresés is egyszerűen módosítható. A szöveget úgynevezett kétrétegű PDF formátumban publikálták, így a felhasználó az eredeti dokumentum képét látja, miközben a keresés a felismert szövegben zajlik.

Az MTI finanszírozásával megvalósult projekt során 800.000 oldalnyi dokumentum feldolgozására került sor, biztosítva ezzel az értékes, ám eddig nehezen felkutatható információk minden állampolgár számára ingyenes elérését. A projekt jelentősége abban rejlik, hogy a 20. századi történelem vonatkozásában kiemelkedő jelentőségű forrásanyagok rossz minőségük ellenére is sikeresen feldolgozásra kerültek, s ezentúl bárki számára bármikor, hozzáférhető a Hírárchívum, melynek szövegei egyedülálló részletességgel, napi bontásban számolnak a 20. század belföldi és külföldi eseményeiről egyaránt.