



„Az elveszettek földje”. Mit őriz meg a *digital preservation*?

Mindenféle hosszú távú archiválásnak lényeges előfeltétele a gondos tervezés. A megfelelő stratégia kialakításához azonban szükség van a „megőrzési szándék” megfogalmazására, vagyis világosan és tárgyilagosan végig kell gondolnunk, hogy mit is akarunk valójában megőrizni. E cikk szerzői az *Ausztrál Nemzeti Könyvtár* számára foglalták össze mindazt, amit ebben a kérdésben érdemes figyelembe venni, de gondolataik hasznosak lehetnek más intézményeknek is, amelyek tudatosabban és hatékonyabban szeretnék digitális archívumaikat gondozni.

Elavulás

Az információhordozók mindegyike idővel fokozatosan tönkremegy, ezért van egy pont, amikor elkerülhetetlenné válik lemásolásuk vagy más médiumra való átköltöztetésük, mielőtt végleg elvesz a rajtuk tárolt tartalom. Nemcsak a környezetből származó fizikai hatások teszik tönkre a tárolóeszközöket, hanem néha a szellemi elavulás veszélye is fennáll, például amikor az illető dokumentum nyelvét már senki nem érti. Ilyen esetekben az objektumnak másolással való megőrzése nem elegendő az információvesztés elkerülésére – át kell menteni valahogy a jelentését is, például magyarázattal vagy fordítással. Bármilyen módszert is használunk a megőrzés során, mindig van némi veszteség: az eredeti objektum egyes elemei nem lesznek reprodukálhatók, mert valamilyen szintig meg kell változtatni bizonyos jellemzőit ahhoz, hogy meg tudjuk oldani a dokumentum túlélését.

Mindezek a digitális dokumentumokra is igazak, hiszen ezeknek is van fizikai hordozójuk (pl. egy merevlemez) és így ezek is ugyanúgy ki vannak téve a külső hatásoknak, mint mondjuk az agyagtáblák vagy a könyvek. De míg egy olyan hagyományos hordozó esetében, mint a papír, a rajta levő információ legalább részben felismerhető mindaddig, amíg porrá nem omlanak a lapok, egy

digitális tároló tartalma általában sokkal korábban használhatatlanná válik, még mielőtt az eszköz teljesen szétesne. Ugyanakkor a digitális információk világában a tárolóeszköz tönkremenetele nem olyan nagy tragédia, mint az analóg dokumentumoknál, hiszen az előbbiek esetében kevésbé tartjuk fontosnak az eredeti példányt, és úgy gondoljuk, hogy a digitális objektumok átvihetők egyik médiumról a másikra anélkül, hogy elvesznének lényeges jellemzőik – még ha ez műszakilag nem is mindig igaz. Mert egy fájlt át tudunk ugyan másolni egy optikai lemezről mágnesszalagra úgy, hogy megmarad a bitek sorrendje, de az a mód, ahogy a két eszközön ezek a nullák és egyesek eltárolódnak, annyira különbözik, hogy ha mikroszkóppal megnézzük, a digitális objektum e két példánya semmilyen hasonlóságot nem mutat.

A számítógépes állományok nagyban függenek olyan egyéb tényezőktől is, mint a hardver és a szoftver. Ha ezek az előfeltételek elérhetetlenné válnak, akkor már jóval azelőtt elvesz a digitális dokumentum értelme/jelentése, még mielőtt kihalna az a nyelv, amelyen írták. Ha például egy könyvtárnak megvan a kopt nyelvű Biblia pergamenen, akkor ez a dokumentum mindaddig jelentést hordoz, amíg akad olyan olvasója, aki érti ezt a nyelvet, és amíg a pergamen tönkre nem megy. Ugyanakkor, ha ez a Biblia digitális másolatban van csak meg, akkor élettartama nemcsak attól függ, hogy vannak-e még kopt nyelven olvasó emberek, hanem attól is, hogy van-e még olyan eszköz a könyvtárban, amellyel ez a fájl megfelelő módon megjeleníthető. A digitális objektumok megőrzése tehát bonyolultabb feladat, mint az analógoké, mert nem elég magát az objektumot tároló fájlt megvédeni a tönkremenéstől, a használatához szükséges környezettel is foglalkozni kell. Ráadásul vannak igen komplex digitális dokumentumok – ezeket a PREMIS 2.0 szótár [1] *összetett objektumnak* vagy *reprezentációnak* nevezi (ilyen lehet pl. egy sok HTML fájlból és képből álló webhely) –, amelyeknél az egyes elemek közötti kapcsolatokat is meg kell valahogyan őrizni, hogy

megmaradjon az eredeti kontextus, amiben azok megjelentek.

Egy mindennapos példával illusztrálva a digitális objektumok használatát: tegyük fel, hogy egy CD-ROM-on levő képet szeretnénk megnézni. Ehhez kell egy megfelelő eszköz, amivel elolvasható az objektum tárolt változata – vagyis egy CD-olvasóval felszerelt számítógép. A CD meghajtó beolvassa a nullák és egyesek sorozatát, tehát az információ fizikailag tárolt formáját, és ezt az egyelőre még jelentés nélküli bitfolyamot átküldi a gép belső memóriájába (RAM), így ehhez más programok is hozzá tudnak férni. Ezután szükség van egy olyan speciális szoftverre (pl. Photoshop), amelyik értelmezi ezt a sorozatot és átalakítja a biteket az egyes képpontokhoz tartozó színek kódokká. Végül a bitfolyamból visszakódolt információt meg kell jeleníteni a felhasználó számára egy képernyőn vagy egy nyomtatón. Fontos megjegyezni, hogy egy digitális objektumot nem feltétlenül csak egyetlen módon lehet prezentálni. Az előbbi példánál maradva: a képfájlt úgy is megjeleníthetjük, hogy a színek helyett a fényerőt és a színtelítettséget adja meg a program minden egyes pixelnél, és ezt a képet is többféle módon lehet közvetíteni a felhasználóhoz.

Nézőpontok

A digitális objektumokat célszerű olyan komplex entitásoknak tekinteni, amelyek több szempontból is vizsgálhatók és kezelhetők. Andrew Wilson 2007-ben készített egy összefoglalást [2] a megőrzésnél figyelembe veendő fontos jellemzőkről. Természetesen az, hogy egy tulajdonság mennyire „fontos”, függ attól a céltől is, amire az objektumot használni akarják, és attól is, hogy annak melyik formájáról beszélünk. Függetlenül attól, hogy egy digitális objektumot az életciklusa különböző fázisaiban egyetlen egységnek vagy részekből összetettnek tekintünk, az mindig különböző formákat vesz fel, annak függvényében, hogy éppen milyen módon használjuk:

- Az objektum *tárolt* alakja a fizikai hordozón való állapotát jelenti, például a kis mélyedéseket az optikai lemezeken, vagy a mágneses töltéseket a szalagon. Előfordulhat olyan eset, hogy egy összetett objektum egyes részei eltérő tárolón vannak, például az internet előtt a BBS-ek (telefonos modemmel hívható faliújságszerű rendszerek) egy részénél az online tartalom részben CD-ROM-on volt, és csak a gyakrabban használt, il-

letve kisebb méretű fájlokat szolgáltatottak merevlemezeiről.

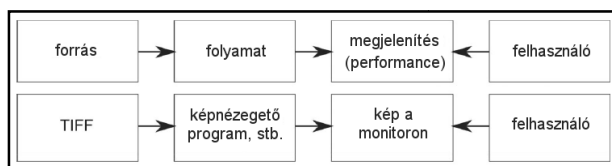
- A *bináris* forma az a bitfolyam, amit a tárolóról egy megfelelő eszköz kiolvas és átvissz a számítógép memóriájába. Tulajdonképpen még ennek is van fizikai hordozója: a RAM-ban levő elektromos töltések. A digitális eszközök egyik nagy előnye, hogy két teljesen eltérő tárolási formából elő lehet állítani ugyanazt a bitsorozatot.
- A bináris állapot *információvá* alakítható át egy megfelelő szoftverrel, vagyis például karaktersorozatokká vagy számértékekké, amelyek jelenthetik, mondjuk egy fénykép készítőjének nevét vagy az egyes képpontok színértékeit. Itt is igaz az, hogy ugyanazt az információt sokféle bináris alakban lehet tárolni. Viszont ha elvész a dekódoló szoftver, akkor általában már nem egyszerű visszanyerni a bitsorozat jelentését.
- A digitális objektumok negyedik állapota maga a *megjelenítés*. Ugyanaz az információtartalom sokféle módon ábrázolható, bár a felhasználók ezek közül rendszerint csak néhányat szoktak találni. Az előbbi fejezetben említett példára visszautalva: a pixelek különböző módokon jeleníthetők meg, sőt a képpontok adatai mellett a fájlban lehetnek szöveges metaadatok is, amelyek szintén megnézhetők egy erre alkalmas szoftverrel.

Megőrzési szempontból más és más fontos jellemzői vannak ennek a négy állapotnak, és ezek tudatosítása segít annak eldöntésében, hogy egy objektum mely tulajdonságait érdemes átmenteni a jövőnek.

Megjelenítés

A digitális objektumok nemcsak többféle nézőpontból vizsgálhatók, hanem a velük való kapcsolat módja is igen változatos lehet. Az ugyan nem gyakori, hogy valaki a tárolt formájukban nézegetné őket (pl. egy lyukkártyán vagy egy optikai lemezen mikroszkóppal), sokkal inkább az általuk hordozott információkra vagyunk kíváncsiak. Ehhez viszont meg kell jeleníteni őket a megfelelő technológiával, és ez a megjelenített alak az, amivel azután kapcsolatba lépünk. Az angol nyelvű szakirodalomban a *representation* szót használják erre, bár eléggé különböző értelemben. Heslop és szerzőtársai 2002-es tanulmányukban [3] a *performance model* fogalmát vezették be, amely magában foglalja az adatforrást (ez lehet pl. egy TIFF fájl), valamint azt a folyamatot, amelynek során a hardver és a szoftver feldolgozza az adatokat, és

egy értelmes, az érzékszerveinkkel felfogható megjelenítést állít elő belőlük (1. ábra).



1. ábra A „performance model”

Ha szeretnénk megfelelő szinten megőrizni egy digitális objektum használhatóságát, akkor fent kell tartanunk azt a mechanizmust, amellyel a jelentést ki lehet nyerni belőle. Ez azonban nem olyan egyértelmű, mint elsőre tűnik. Hogy mi a „megfelelő szint”, az nagyban függ attól, hogy mire akarjuk ezt a dokumentumot használni. Egy fotó esetében például elegendő lehet önmagában megjeleníteni az azt tároló TIFF állományt, de ha egy olyan TIFF fájlról van szó, ami valójában egy 3D-s objektum egyik apró eleme (pl. egyik mintázata), akkor nem elég csak azt megmutatni a nézőnek. Ezen kívül a felhasználó igényei is befolyásolják, hogy mi a számára „megfelelő”. Például egy hangfelvétel esetében egyesek meglepészenek azzal a minőséggel, ami a gépre kapcsolt asztali hangszórókból hallható, de egy hangtechnikus azokat a frekvenciákat is hallani szeretné, amelyeket ezek a kis dobozok nem tudnak reprodukálni. Tovább bonyolítja a helyzetet, hogy a felhasználó nem mindig csak passzív befogadója az információnak, hanem maga is befolyásolhatja a megjelenítést (pl. interaktív digitális műalkotásoknál). Az eredetivel tökéletesen azonos reprodukálást csak úgy lehetne elérni, ha fenntartanánk az eredeti berendezéseket és szoftvereket az idők végezetéig, ami nyilván irreális elvárás. A *performance model* értelmezése szerint azonban a feladat nem megoldhatatlan, mert nem fontos sem az adatforrást, sem a megjelenítéséhez szükséges folyamatot eredeti állapotában megőrizni, csak az a lényeg, hogy ugyanazt az „előadást” kapjuk vissza, mint ami az eredeti volt. Vagyis a modell egyes elemei megváltoztathatók, kicserélhetőek, és mindaddig, amíg lényeges változás nem történik a megjelenített objektum alapvető jellemzőiben, azt hitelesnek lehet tekinteni. Persze az meglehetősen szubjektív, hogy mit értünk a „hiteles” és az „alapvető” szavakon, és bár a digitális dokumentumoknál valóban nincs akkora jelentősége az eredeti példánynak, mint a papíralapúaknál, az objektum eredeti állapotának mégiscsak lehet olyan fajta „hitelességgel”, ami miatt érdemes azt megtartani, még ha a megjelenítési folyamaton változtatunk is az idők során. Ilyen

esetekben tehát a forrás megváltoztatása nem elfogadható megoldás. Az egyik formátumról a másikra való átköltöztetése (migrálása) felveti annak a veszélyét is, hogy bár az így átkódolt állomány továbbra is alkalmas lehet egy bizonyos megjelenítés elfogadható pontosságú reprodukálására, de megtörténhet, hogy egy másfajta megjelenítésre viszont már nem alkalmas, amely pedig egyesek számára ugyanolyan fontos, vagy akár még fontosabb is lenne.

A megőrzési szándék

Hogy mit tekintünk „megfelelő” szintű megőrzésnek, az relatív, és a szakirodalomban nagy viták folynak erről. Attól függően, hogy az adatforrás mely jellemzőit tartjuk fontosabbnak vagy megváltoztathatónak, és hogy a megjelenítés mely formáit tekintjük elfogadhatóbbnak, a különböző megőrzési technológiák többé vagy kevésbé bizonyulnak megfelelőnek. Ha egy erre szakosodott intézmény nem határozza meg világosan, hogy a digitális objektumoknak milyen tulajdonságait kívánja megőrizni a jövő számára, akkor nem nagyon tud helyes és hatékony stratégiát kidolgozni, és nehéz lesz később a megőrzési tevékenység eredményességének megítélése is. Ha például fontos megtartani a digitális objektum bizonyos részeit eredeti állapotukban, akkor a migráció nem alkalmas technológia, helyette a megjelenítő környezet emulálását kell választani, amihez nem szükséges megváltoztatni az adatforrást. Ha nincs ilyen elvárás, akkor a várható felhasználás függvényében kell döntenie a különböző migrálási módszerek közül: például, ha egy Excel táblázatot PDF-re konvertálunk, akkor megőrizzük a külalakját, de elveszítjük a beleépített képleteket; ha viszont a nyílt ODF formátumra alakítjuk át, akkor lehetnek ugyan némi változások a külalakjában, de működőképesek maradnak a képleteket tartalmazó cellák. Az első esetben a megjelenítés helyességét ítéltük fontosabbnak, a másodiknál a fájlban kódolt információ teljes körű megőrzését.

Mielőtt belekezdenénk a megőrzési folyamatba, vagy akár a stratégia kidolgozásába, meg kell határoznunk, hogy mit is szándékozunk elérni? A „megőrzési szándék” magában foglalja azt az elhatározást, hogy meg kívánunk őrizni bizonyos digitális objektumokat, továbbá a megőrzés körülményeit és céljait, valamint annak eldöntését, hogy meddig akarjuk archiválni őket. A következőket kell tehát mindenképp előtte megtenni:

- Annak megállapítása, hogy igény vagy szándék van bizonyos objektumok megőrzésére.
- A megőrzendő objektumok jellemzőinek megismerése. (Ehhez szükség lehet az objektumok különböző szinteken történő leírására, általában annak függvényében, hogy az intézmény milyen szinten szeretne foglalkozni velük, és milyen fokú kontextuális, illetve technikai ismeretei vannak róluk.)
- Az időtáv definiálása. (Ez lehet egészen rövid: pl. néhány nap vagy hónap, de lehet „örökké” is.)

Egy hosszabb példával illusztrálva a különböző célokat és a belőlük következő optimális megőrzési technológiákat, térjünk vissza a CD-ROM-on tárolt képhez. Lehet, hogy egyes intézményekben csak az a fontos, hogy ez a kép könnyen megnézhető legyen bármikor és bárhol, lehetőleg minél kevesebb minőségváltozással. Vagyis itt egyfajta megjelenítés élvez elsőbbséget a fájl tárolt, bináris és információs formáival szemben. Ez esetben ésszerű úgy dönteni, hogy a fájlt TIFF formátumra migrálva archiváljuk egy háttértáron és a mindennapi használatra – némi minőségromlással és bizonyos beágyazott metaadatok elvesztésével – egy kisebb JPG verziót is konvertálunk belőle. Más gyűjteményeknél viszont lehet, hogy fontos a kép metaadatainak megőrzése is a fájlban (pl. a fotós neve, a kamera típusa stb.), továbbá a kép olyan jellemzőinek megtartása, mint a színtér, a képméret stb. Ilyenkor olyan migrációs formátumokat szabad csak használni, amelyeknél ezek az információk nem vesznek el. Végül olyan archívumok is létezhetnek, amelyeknél a fájl eredeti bináris állapotát és azt a CD-ROM-ot is meg kívánják tartani, amelyen beérkezett. Ebben az esetben is célszerű lehet egy TIFF változat előállítását időnkénti megtekintés céljából, de vigyázni kell a hordozóra is (pl. az élettartamot maximalizálható körülmények között kell tárolni az optikai lemezt), és meg kell tartani az eredeti hardvert (pótalkatrésszel) és szoftvert is, amivel az szükség esetén lejátszható. A valóságban az is előfordulhat, hogy ugyanazon intézményen belül ugyanazzal a digitális

objektummal kapcsolatban különböző megőrzési szándékok léteznek, attól függően, hogy milyen környezetben, milyen céllal akarják azt majd használni egyesek.

A szerzők tapasztalata szerint, ha nincsenek világos előírások arra, hogy mit és hogyan kell megőrizni egy archiváló intézményben, akkor a megőrzési folyamat auditálása igencsak frusztráló és kis határfokú tevékenység. Ha viszont világosan megfogalmazták a megőrzésre vonatkozó szabályokat és azt is, hogy melyek a még elfogadható változtatások, akkor ez sokat segít abban, hogy csak olyan módosítások történhessenek a jövőben az archivált objektumokon, amelyek tervezettek és megjósolható következményekkel járnak. Ez alapvetően fontos ahhoz, hogy megbízható és számon kérhető megőrzési stratégiákat lehessen megvalósítani a digitális megőrzéssel foglalkozó nagyméretű projekteknél.

Hivatkozások

- [1] PREMIS Editorial Committee: PREMIS 2.0 Dictionary. 2008.
<http://www.loc.gov/standards/premis/v2/premis-2-0.pdf> (letöltve: 2010.01.21.)
- [2] WILSON, A.: Significant properties report (InSPECT Work Package 2.2). Arts and Humanities Data Service, 2007.
http://www.significantproperties.org.uk/wp22_significant_properties.pdf (letöltve: 2010.01.21.)
- [3] HESLOP, H. – DAVIS, S. – WILSON, A.: An approach to the preservation of digital records, 2002. = http://www.naa.gov.au/Images/An-approach-Green-Paper_tcm2-888.pdf (letöltve: 2010.01.21.)

/DEL POZO, Nick – LONG, Andrew Stawowczyk – PEARSON, David: „Land of the lost”: a discussion of what can be preserved through digital preservation. = *Library Hi Tech*, 28. köt. 2. sz. 2010. p. 290–300/

(Drótos László)