

**Körmendy Lajos**

***Elektronikus iratok és levéltár***

**Körmendy Lajos**

# **Elektronikus iratok és levéltár**

A kiadásért felel: Körmendy Lajos  
ISBN 978-615-00-0831-8

© Körmendy Lajos 2017

## Tartalom

Előszó	5
1 A levéltárak funkciói és az iratok értékei	7
1.1 A levéltárak funkciói	7
1.2 Az iratok értékei	8
1.2.1 Az adminisztratív/gyakorlati érték	8
1.2.2 A jogbiztosító érték	9
1.2.3 A történeti érték	10
2 Levéltári kihívások az analóg és a digitális világban	12
2.1 Kihívások és levéltári válaszok	12
2.2 Informatika, kommunikáció és elektronikus iratok	14
2.3 Az emberi emlékezet digitalizálása és adatrobbanás	16
2.4 Az elektronikus iratok tartós megőrzésének nehézségei	17
3 A levéltári irat és alkotó elemei	19
3.1 Hordozó	19
3.2 Forma	20
3.3 Tartalom	21
3.4 Azonosító	21
3.5 Struktúra	22
3.5.1 Belső struktúra	22
3.5.2 Külső struktúra	23
3.6 Kontextus	24
3.7 Az elektronikus iratok alkotó elemeinek szétválása	25
3.8 A levéltári iratok egyedi jellege	26
4 A metaadatok	27
5 Az iratok integritása és lényeges jellemzői	30
6 Az iratok hitelessége	34
7 A fájlok (megőrzési) formátumai	37
8 A szabványok	42
8.1 XML	44
8.2 MOREQ	46
8.3 OAIS	47
8.4 ISAD(g) és SAAR(CPF)	47
8.5 EAD és EAC	49
9 Az elektronikus iratok tartós megőrzésének stratégiái	51

9.1	Migráció	54
9.2	Emuláció	56
9.3	Analóg és hibrid megőrzés	57
9.4	Különleges iratok	59
9.4.1	E-mail-ek	59
9.4.2	Adatbázisok	61
9.4.3	Weblapok	63
10	Az elektronikus iratok fizikai hordozóinak megőrzése	65
11	A levéltárban átalakulása az elektronikus iratok korában	68
12	Az elektronikus iratok megőrzésének tétje	71

## Előszó

Mintegy négy évtizedes (1975-2012) levéltári karrierem alatt szemtanúja és résztvevője lehettem annak a folyamatnak, amely során az informatika átalakította az iratokat, a levéltártant és természetesen magukat a levéltárakat is – teljesen logikusan, hiszen maga a modern társadalom is átalakult az információs-kommunikációs forradalom alatt. 25 évig foglalkoztam levéltári informatikai kérdésekkel, a gondolataimat kéttucatnyi publikációban osztottam meg az érdeklődő kollégákkal. Nyugdíjba vonulásom után úgy éreztem, hogy az elektronikus iratokkal kapcsolatos tapasztalataimat és tudásomat érdemes egy összefoglaló műben összegezni. Ösztönzött erre az is, hogy nem voltam teljesen elégedett a 2009-ben megjelent Levéltári kézikönyv vonatkozó fejezetével, és nemcsak az óhatatlanul bekövetkező gyors elavulás miatt. (A kiadványnak én voltam a főszerkesztője, és a kérdéses fejezet nagy részét is én írtam.) Ezért elhatároztam, hogy „újraírom” ezt a részt, más szemszögből, más megközelítéssel, de felhasználva a korábbi munkáimat. Az írás alatt három fő célkitűzés vezérelt: 1) összefoglaló képet adni az e-iratok problematikájáról; 2) megvilágítani a legmélyebb archivisztikai összefüggéseket; 3) az egész kérdést történeti és társadalmi keretbe helyezni.

Az e-iratokkal foglalkozó szakirodalom hatalmas, már meglehetősen előrehaladt a szakosodás, ezért kevesen képesek átlátni az egész kérdéskört a maga bonyolult összefüggéseivel. Emiatt sok levéltáros „csinálja, bár nem érti”, és joggal feltételezhetjük, hogyha majd értik (az egészet), akkor jobban fogják csinálni (a maguk részét). De nem elég megérteni a makroösszefüggéseket, ismerni kell a mikrovilágot is, mert bármilyen furcsának tűnik is, az utóbbi sok tekintetben meghatározza az előbbit. (Hasonlóképpen, mint a fizikában: az atomi és a szubatomi szinten érvényes törvények határozták/határozzák meg a csillagok, a bolygórendszerek és a galaxisok, sőt az egész Világegyetem születését, működését és halálát.) Erre tettem kísérletet különösen a 3. fejezetben, ahol igyekeztem az e-irat elemi összetevőit definiálni és a működésüket röviden leírni.

Talán a harmadik célkitűzés megvalósítása, a történeti és társadalmi keret fölvezetése jelentette számomra a legnagyobb kihívást. Ezzel nem nagyon foglalkoznak a teoretikusok, pedig alapvetően fontos szakmafilozófiai kérdés, hogy a levéltári elektronikus korszakot (benne az e-iratokat) különálló újként kezeljük-e vagy pedig kontinuumként. Én feltétlenül a második opció mellett teszem le a voksot, és remélem, hogy ebben a műben sikerült ezt bebizonyítanom. Ezért hívom fel mindig a figyelmet az analóg és az e-világ közötti azonosságra vagy hasonlóságra, persze nem hallgatva el a különbözőségeket sem. A tár-

sadalmi vonatkozások azért fontosak, mert a levéltárak a társadalom részei, az ott fölmerülő kérdésekre kell reflektálniuk, annál is inkább, mert a társadalom határozza meg a levéltárak működését.

Az informatika hihetetlen gyorsasággal fejlődik, ezért minden vele kapcsolatos mű gyorsan avul. Remélem, hogy ez a megállapítás az írásomra csak részben lesz érvényes, egyrészt, mert a levéltártani részek sokkal időtállóbbak, másrészt az informatikában is az állandóra vagy a lassan változó lényegre helyeztem a hangsúlyt. A kézirat a 2015. évi állapotot írja le.

Budapest, 2017 november

# 1 A levéltárak funkciói és az iratok értékei

## 1.1 A levéltárak funkciói

Az iratokat mindig meghatározott céllal gyűjtötték és őrizték, ennek megfelelően a levéltárak (legyenek azok irattárak vagy történeti gyűjtemények) négy fő funkcióját különböztetjük meg: az igazgatásit, a jogbiztosítót, a forrásbázist, és a kulturálist.

- *Igazgatás.* A szervezett emberi társadalmak működéséhez egy adott fejlettségi szint fölött már elengedhetetlen az írásbeli rögzítés, mert a szükséges adatok nyilvántartása nélkül nem lehetne igazgatni, szervezni, elbírálni egyének és csoportok, hivatalok és szervezetek (együtt)működését. A hivatalok és szervezetek levéltárai elsősorban azzal a céllal jönnek létre, hogy a saját működésüket szolgálják, de akár egyének is felhalmozhatnak iratokat hasonló céllal.
- *Jogbiztosítás.* A fejlett, írásbeliséggel bíró társadalmak kiépülésével a jogbiztonságot szolgálta, hogy az elnyert jogokat (pl. nemesi rangot, birtokadományt) írásban rögzítették, amit aztán erre hivatott intézmények megőriztek, hogy a későbbiekben ezekkel az iratokkal lehessen bizonyítani vagy cáfolni a jogokat. Ilyen volt pl. az antik Rómában a Tabularium, amely magániratok depositja volt, vagy a magyar királyi kancellária levéltára, ahol a Libri regii c. sorozatban tartották nyilván a királyi cím- és rangadományokat. A 19. századtól kiépülő polgári társadalmakban a jogbiztonság alapvető követelménnyé vált, és ennek megfelelően újabb, a jogbiztonságot szolgáló levéltárak jöttek létre, pl. a közjegyzői levéltárak, az állami anyakönyvek gyűjteményei, a telekkönyvi nyilvántartások. Jogállamban egy hiteles körülmények között keletkezett és egy hiteles őrző intézményben, pl. állami levéltárban megfelelően archivált iratok bármelyike felhasználható bizonyítékként.
- *Forrásbázis.* A levéltári adatokat nemcsak igazgatási vagy jogérvényesítő céllal használhatják, hanem tudományos és/vagy a múlt megismerésének céljából is. A legismertebb ilyen felhasználás a történeti kutatás, amikor már „elavult”, azaz gyakorlati célra már nem szolgáló, levéltárakban őrzött iratokat használnak forrásként. De hasonló funkcióval bírnak azok a hatalmas, archivált adategyüttesek is, pl. geológiai vagy csillagászati mérések adatbázisai, amelyeket későbbi kutatók használnak fel tudományos munkájukban.

- A *kulturális* funkció csak viszonylag későn, a 20. században jelent meg a levéltárak életében. A levéltári kiállítások, látogatások, ismeretterjesztő előadások és publikációk a levéltárról magáról vagy a benne őrzött iratokról azt a célt szolgálják, hogy a szélesebb közönséggel megismertessék a levéltár munkáját és feladatait, valamint hogy az iratok értékére felhívják a figyelmet.

A levéltári funkciók szoros kapcsolatban vannak az őrzött iratok értékével, amelyeket mindig az adott társadalom determinál. Amint a következő fejezetben meglátjuk, a történelem során akár rövid idő alatt is radikális változások történhettek a (levéltári) iratok értékeiben, amelyek így maguk után vonják a funkciók érvényesülésének változását is.

## 1.2 Az iratok értékei<sup>1</sup>

Az iratok az életciklusuk során<sup>2</sup> – létrehozásuk pillanatától a levéltárba adásukig, illetve kiselejtezésükig – többféle értéket hordoznak. Ezek az értékek egyidejűleg vannak jelen és súlyuk az idők folyamán állandóan változik.

### 1.2.1 Az adminisztratív/gyakorlati érték

Az adminisztratív/használati érték az a (potenciális) előny vagy haszon, amit a használó élvez az iratok információja révén a jelen vagy jövőbeni tevékenységében (nem számítva a jogbiztosító tevékenységet). Az iratképző számára ez a legfontosabb érték, hiszen az iratokat éppen azért hozta létre, hogy a tevékenységét segítsék. Ahogy azonban lezárulnak az ügyek, a hozzájuk tartozó iratok adminisztratív/használati értéke csökken, és mivel a cselekvés, tevékenység időben zajlik, az értékcsökkenés kapcsolatban áll az idővel. Az egyes iratok értéke jelentősen eltérhet, az értékállandóságuk úgyszintén. Egy üzlet kis értékű megrendelésről szóló ügyirata a teljesítés után rohamosan veszít az adminisztratív/használati értékéből, a cég 1-2 éven belül valószínűleg ki is selejtezi. A Coca-Cola üdítőital-koncentrátum titkos receptjének használati értéke – az irat több mint 140 éves, és újsághírek szerint csak a cég néhány kiválasztott dolgozója ismeri – viszont az idők folyamán nem hogy csökkent volna, hanem ellenkezőleg, nőtt, mert a cég folyamatosan terjeszkedett és növelte a forgalmát. Az ilyen tendencia azonban kivétel, az iratok túlnyomó

<sup>1</sup> Ennek a fejezetnek a részletesebb kifejtését lásd A levéltári irat értéke és az iratértékelés című, a *Levéltári Szemle*, 1994/4. számában megjelent cikkemben.

<sup>2</sup> A levéltári szakirodalomban általánosan elfogadott nézet szerint az iratoknak létezésük alatt három ciklusuk lehet: az aktív (a keletkezésük és aktív használatuk a létrehozónál), a félig aktív (amikor a létrehozónál az aktív használatból kikerültek, de még nem selejtezték ki, illetve nem szállították őket levéltárba), végül az inaktív (amikor már a levéltárba kerültek végső megőrzésre). A kiselejtezett iratoknál a három ciklus egyre vagy kettőre rövidülhet. Lásd pl. *Terminologie archivistique multilingue de l'ICA*. Elérhető: <http://www.ciscra.org/mat/termdb/term/1028>

részére a használati érték gyors csökkenése a jellemző. Mindazonáltal soha sem lehet azt mondani, hogy egy, akár évszázadok óta nem használt irat értéke is nulla lenne, mert bármikor történhet olyan esemény, amihez fel lehet használni. Sok olyan eset történt, hogy pl. hosszú idő óta „elfeledett”, levéltárakban őrzött tervrajzokat felhasználták épületek rekonstrukciójához.<sup>3</sup>

### 1.2.2 A jogbiztosító érték

Jogbiztosító érték az a tényleges vagy potenciális előny vagy haszon, amit a használó élvez az iratok információja révén jogi státusza vagy jogi vonatkozású állítása törvényes bizonyításakor. Ez is alapvető fontosságú az iratképzőnek, bár nyilvánvaló, hogy sokkal kevesebb iratot hoznak létre jogbiztosító, mint használati szándékkal, mert az iratot az iratképző ellen is fel lehet használni bizonyítékként, aki számára így negatív értékű lesz. (Ha ezt sejtette és tehetné volna, biztosan nem hozta volna azt létre.) A jogrendnek megfelelően a jogok is elavulnak, semmissé lesznek, egyesek gyorsabban, mások lassabban. Vannak olyan iratok, amelyek percek alatt elveszítik bizonyító erejüket, illetve annak egy részét: pl. egy pénztárblokkal nagy valószínűséggel hiába reklamálunk egy szupermarketben, hogy kevés visszajáró pénzt kaptunk tíz perccel ezelőtt,<sup>4</sup> vannak viszont tartós értékű iratok, pl. a telekkönyvek. Az esetek többségében írott törvények határozzák meg, hogy egy-egy, azonos típusba tartozó ügyek (pl. adóügyek) elévülési ideje mennyi, tehát az értékvesztés egyes iratfajtáknál egyidejű és tömeges. Az iratok összességére az idők folyamán az érték csökkenése a jellemző – hasonlóan a kurrens használati értékhez –, mint ahogy itt sem lehet azt mondani, hogy egy, akár évszázados irat jogbiztosító értéke nulla lenne, mert az iratok potenciálisan mindig felhasználhatók jogi bizonyításra. Vannak kuriózumok a jogtörténetben, mint pl. az a pár évtizede még az NSZK-ban lezajlott per Lübeck és Mecklenburg között, ahol a bíróság bizonyítékként elfogadott egy 13. századi iratot.<sup>5</sup> Ennél jobb példa, mert tipikusabb és tömeges, az, ami Kelet-Európában a 20. század második felében történt. Ismeretes, hogy 1945 után, de különösen a kommunista hatalomátvételt követően a régióban tömeges, kártalanítás nélküli államosítások történtek. Egy-egy államosítási törvény vagy rendelet hatálybalépésének napján tulajdoni lapok, adásvételi szerződések és egyéb, tulajdonnal kapcsolatos iratok százezrei veszítették el bizonyító erejüket. Az 1989-es fordulat után a hatalomra jutott demokratikus kormányok jogilag is

---

<sup>3</sup> Lásd pl. *Budapest enciklopédia* (Budapest 1981) 70. old.

<sup>4</sup> De ugyanez a számla egy bűnügyben akár évek múlva is bizonyítékként szolgálhat.

<sup>5</sup> Carl Haase: Studien zum Kassationsproblem. In: *Der Archivar*, November 1975, 408. old.

igyekeztek „rehabilitálni” a magántulajdont,<sup>6</sup> és ha valaki bizonyítani tudta, hogy elkobozták a vagyonát, akkor pénzbeli kárpótlásra jogosult. Tehát az előbb említett tulajdoni lapok, adásvételi szerződések több évtized után – már ha megmaradtak – újra fontos bizonyító dokumentumokká váltak.

### 1.2.3A történeti érték

Történeti érték az a fontosság, amit a használó tulajdonít az irat információjának a múlt megismerése kapcsán. Ez az érték az iratképzőnek tulajdonképpen közömbös, ő az iratokat használati vagy jogbiztosító céllal hozta létre. Az idő múlásával fokozatosan csökken az egy adott időszakban keletkezett iratok mennyisége selejtezések és pusztulások miatt. Ezzel párhuzamosan halványul az egyéni és a kollektív emberi emlékezet, tehát az adott korról szerezhető információs források állandóan szűkülnek. A ritkaság pedig értéknövelő hatású, mind szellemi, mind anyagi szempontból. Ennek következtében az előző két értékkel ellentétben a történeti érték időskiban ábrázolva a nulla közeléből indul, és az idő múlásával fokozatosan emelkedik, azaz a megmaradó iratok egyre fontosabbak lesznek.

Viszont a történeti érték sokkal több szubjektív elemet takar, mint a másik kettő, hiszen a jogbiztosító értéket tulajdonképpen jogszabályok határozzák meg, az adminisztratív/használati érték pedig többnyire kiszámítható, rövidtávú tevékenységhez kötődik. A szubjektív elemek ellenére mégis vannak olyan külső tényezők, amelyek - társadalmi szinten - determinálják az aktuális történeti értéket.

- *A társadalmi értékrend.* Egy példa: az USA-ban az ötvenes években indult egy spontán mozgalom az ősök felkutatására, aminek eredményeként egy-két évtized alatt Észak-Amerikában, majd Európában egyszerű emberek (tehát nem nemesi származásúak) tömegei fedezték fel a levéltári forrásokat, különösen az anyakönyveket és az összeírásokat. Az ezredforduló táján már több országban ők tették ki a kutatók többségét.<sup>7</sup> Mindez egy társadalmi változás, egy identitáskeresés és tudati demokratizálódás eredménye volt, ami - és ez hangsúlyozandó - lényegében független volt a politikától, mert még a kelet-európai, sokáig diktatórikus országokban is hasonló volt a tendencia.
- *A politika.* A kelet-európai országokban a 20. század második felében a vulgármarxista ideológia vált hivatalossá. Központilag elrendelt, a politikának tetsző

---

<sup>6</sup> Például Magyarországon az ún kárpótlási (XXV/1991. sz.) törvény.

<sup>7</sup> Lásd pl. Gérard Ermisse: La communication. In: *La gazette des archives*, 1988 No. 141, 200. old., és L'étude sur les publics des Archives de France. In: *Comma*, 2003/2-3. sz., 71. old.

kutatási programok indultak el, történészek százai kezdtek el foglalkozni gazdaságtörténeti kérdésekkel, vizsgálták a parasztfelkelések és a munkásmozgalom történetét - elég csak megnézni az 1950/1960-as években kiadott történeti művek címeit. Eltekintve a kutatási eredmények minőségétől - volt közöttük jó és rossz is -, kétségbevonhatatlan, hogy ezek a témák a történelem addig elhanyagolt területeit szólaltatták meg, viszont a politikai kényszer állította őket reflektorfénybe. Az adott kor kutatói számára a témára vonatkozó források mindenesetre jelentősen felértékelődtek.

- *A tudomány.* Ha tanulmányozzuk a historiográfia történetét, megállapíthatjuk, hogy a történeti kutatás az idő haladtával újabb és újabb területekre terjedt ki. Kezdetben csak királyok és hadvezérek dicső tetteinek leírásával foglalkozott, majd fokozatosan vonta vizsgálat alá a gazdaság, a társadalom, a kultúra további területeit. Ki beszélt 100 évvel ezelőtt mentalitástörténetről vagy klímátörténetről vagy 50 évvel ezelőtt nő-történetről? Vannak „divatok” is, amelyek egy-egy téma intenzív kutatását eredményezik, a divat lényegének megfelelően viszonylag rövid időre. Az új területek bevonása viszont új, addig figyelmen kívül hagyott források felfedezését jelentette és jelenti, amelyeknek az érdeklődés nyomán megnő az értékük. Aztán, ha kikerülnek az érdeklődés középpontjából, akkor csökken az értékük. Nem tudhatjuk, hogy a történettudomány „tágulása”, vagy „osztódása” meddig fog tartani, valószínűleg a jövőben is fog találni olyan részterületeket, amelyek eddig nem kerültek előtérbe.

\* \* \*

Fölmerülhet a kérdés, hogy a levéltárak forrásbázis funkciójánál említett tudományos adategyüttesek, pl. geológiai vagy csillagászati mérések adatbázisai mely érték kategóriába sorolhatók. Nos, tartozhatnak akár az adminisztratív/használati értékhez, akár a történeti értékhez, ez szemlélet kérdése.

Az értékelemzés végére érvén levonhatunk néhány következtetést. Az első, hogy amíg az irat létezik, a három érték (az adminisztratív/gyakorlati, a bizonyító és a történeti) állandóan jelen van, és együttesen határozza meg az irat információs értékét. A másik, hogy az idők során a történeti érték fokozatosan kiszorítja a másik kettőt. A harmadik, hogy bár az egyes iratok vagy iratcsoportok esetében nagy az értékek instabilitása, az iratok összességére vonatkozó le- vagy felértékelődési trendek ettől függetlenül érvényesülnek.

## 2 Levéltári kihívások az analóg és a digitális világban

### 2.1 Kihívások és levéltári válaszok<sup>8</sup>

Visszatekintve a levéltárak történetének elmúlt 130 évére érdekes szakaszosságot figyelhetünk meg a szakma fejlődésében: 50-60 évenként történtek olyan minőségi változások, amelyek a levéltárban lényegi részeit vagy egyenesen alapjait érintik. Természetesen nem egyik napról vagy évről a másikra következtek be az „ugrások”, azokat megelőzte egy-egy 2-3 évtizedes időszak, amit a „problémák halmozódása” kifejezéssel jellemezhetünk: az új jelenségek először szórványosan jelentkeztek, majd egyre inkább terjedtek, a periódus végén pedig már feszítő gondot okoztak. A feszültségek aztán a szakma megújulásával enyhültek, majd minden kezdődött előlről, de már más feltételrendszerben és összefüggésben.

A 19. század végén az európai polgári társadalmak fejlődésével került napirendre a szervezett modern levéltárak létrehozásának szükségessége: az állam működése, de maga a társadalom is megkövetelte, hogy a köziratok – és ettől nem függetlenül a magániratok – megőrződjenek és visszakereshetők legyenek. A szakmai közösség legfontosabb válasza a kihívásra a provenienciaelv megfogalmazása volt, ami lehetővé tette, hogy a levéltárak átfogóan begyűjtsék és feldolgozzák a szervek, szervezetek és jelentős magánszemélyek iratait.

1898-ban jelent meg Samuel Muller, Johan Feith és Robert Fruin holland levéltárosok kézikönyve, amit sokan a modern levéltártudomány kezdetének tekintenek.<sup>9</sup> A mű, amelynek alapelvei ma is helytállóak, a fond koncepciójára épült, melynek lényege benne rejlik a fond definíciójában: egy jogi vagy természetes személy rendeltetésszerű működése során képzett iratok együttese. A provenienciaelv sarokköve a fondok tiszteletének az elve, azaz hogy egy fondba tartozó iratokat nem szabad keverni más fondba tartozókkal. A provenienciaelvet sokféleképpen interpretálják, de alapvetően két felfogást különböztethetünk meg:<sup>10</sup>

---

<sup>8</sup> E fejezet témájának részletesebb kifejtését lásd Körmeny Lajos (szerk.): *Levéltári kézikönyv* (Osiris Kiadó 2009) 467-472. old.

<sup>9</sup> Samuel Muller & Johan A. Feith & Robert Fruin: *Manual for the Arrangement and Development of Archives* (New York 1940 - angol fordítás)

<sup>10</sup> Az általam ismert szakirodalomban a különböző interpretációkra a legjobb áttekintést a *The Principle of Provenance, First Stockholm Conference on Archival Theory and the Principle of Provenance, 2-3 September 1993* (Skrifter utgivna av Svenska Riksarkivet 10, Stockholm 1994) c. kiadványban találtam.

- Az egyszerűbb, kevésbé szigorú felfogás csak a fondok tisztelete elvét tekinti fontosnak, ami azt jelenti, hogy a levéltárban egy fond iratai nem keveredhetnek egy másik fond irataival, viszont a fondon belüli struktúra kialakításában nagy szabadságot ad a levéltárosnak.
- A szigorúbb felfogás a fondok tiszteletén túl respektálja a fondokon belüli eredeti rendet is, mondván, hogy az lényegében tükrözi a létrehozó szerv struktúráját; ez azt eredményezi, hogy levéltári feldolgozáskor az iratképző által kialakított iratstruktúrát a lehetőség szerint nem változtatják.<sup>11</sup>

A provenienciaelv alkalmazása előtt a levéltárosok igyekeztek darabszinten leírni az iratok tartalmát, kiemelni belőlük a kulcsszavakat (mutatók) vagy tárgyi, formai csoportokba rendezték az iratokat, szintén ellátva őket különböző segédletekkel (pl. listákkal, mutatókkal). A bürokrácia által termelt, gyorsan növekvő iratmennyiség már az újkorban lehetetlenné tette a darabszintű leírást, a mesterséges tárgyi besorolás pedig, mivel rendkívül szubjektív, nem tette lehetővé a megfelelő információ-visszakeresést. A besorolás ugyanis mindig az indexáló érdeklődését és tudását tükrözi, a kutató, különösen egy késői korban, már más fogalmakban gondolkodik, így nagy valószínűséggel nem fogja megtalálni az általa keresett információt. Elég megnézni a 18. vagy 19. századi indexeket, hogy rádöbbenjünk, mennyire más tárgyszavakat tartottak fontosnak az adott kor levéltárosai, mint mi. Ráadásul a jogállamiság kiépülésével kétségessé vált az ilyen drasztikusan átrendezett, kiselejtezt iratok hitelessége is. A provenienciaelv alkalmazásával viszont a hitelesség jól megőrizhető, és az információk is jó eséllyel megtalálhatók. (Ennek részletezését lásd a 3.8. fejezetben.)<sup>12</sup>

A következő nagy változás a levéltártanban a 20. század közepére datálódik, amikor elsősorban az állam egyre kiterjedtebb szerepvállalása miatt ugrásszerűen megnőtt a keletkezett és feldolgozandó iratok mennyisége. Itt az új korszak nem köthető egy olyan ki-

---

<sup>11</sup> Elsősorban az adott ország adminisztrációs és irattári gyakorlatától, illetve hagyományaitól függ, hogy hogyan fedi az iratok struktúrája az iratképző szervezeti felépítését. Ahol a német eredetű ún. regisztratúra rendszer (irattári terv) hagyománya érvényesül, a megfelelés általános, másutt kevésbé. Lásd Michel Duchein: Le respect des fonds en archivistique: principes théoriques et problèmes pratique. In: *La gazette des archives*, 1977, 90-91. old., valamint a Le principe de provenance et la pratique du tri, du classement et de la description en archivistique contemporaine. In: *Janus* 1998.1, 96-97. old., valamint Peter Horsman: The Last Dance of the Phoenix or the De-discovery of the Archival Fonds. In: *Archivaria* No. 54, Fall 2002, 12. old.

<sup>12</sup> A provenienciaelvnek komoly hátrányai is vannak. Az adminisztráció, főképp az állami szektorban, gyorsan és gyakran változik: szervek, hivatalok szűnnek meg, olvadnak be más szervekbe, újak jönnek létre, és ezek a változások gyakran régi fondok lezárását és újak születését eredményezik. Egy adott ügykör – pl. bevándorlási ügyek – kompetenciája akár fétucatnyi hivatal között is vándorolhat 20-25 év alatt, ami levéltári szempontból nem hosszú idő, és ez nagyon megnehezíti a kutató feladatát. Eme hiányosság orvoslására dolgozták ki ausztrál levéltárosok az ún. sorozat elvet, aminek kifejtésére itt nincs hely. (Erről lásd pl. Peter Scott: The Record Group Concept: a Case of Abandonment. In: *American Archivist* 29 (1966) 493-505. old.

emelkedő új szakmai elvhez, mint amilyen a provenienciaelv volt, sokkal inkább új módszerek sorozatát konstatálhatjuk az iratértékelésben, a selejtezésben (például mintavétel), a mikrofilmmezésben (ennek általánossá válása is ekkor következik be) és az állományvédelemben (tömeges konzerválás, savtalanítás stb). A fentiek közül érdemes részletesebben kitérni az iratértékelésre.

Az USA-ban először az 1930-as években, majd különösen a következő évtizedekben szembesültek az iratok mennyiségének ugrásszerű növekedésével, ami a nagyobb állami szerep (New Deal, világháború) következménye volt. Az új iratértékelési koncepciót, ami lehetővé tette az értékes iratok hiteles megőrzését, ugyanakkor a nagyobb selejtezést is, Theodore R. Schellenberg, a National Archives levéltárosa dolgozta ki. A levéltárosnak, ha nem akarta, hogy elborítsa az irattömeg, a szervnél működő „iratmenedzserrel” (records manager) együttműködve kellett értékelni, selejtezni, előkészíteni az iratokat, még a levéltárba szállítás előtt. Még az értékelés előtt meg kell ismernie a szerv helyét az adminisztratív hierarchiában, tisztában kell lennie a szerv egészének és szervezeti egységeinek pontos funkcióival, tevékenységükkel és működésükkel, valamint azzal a szélesebb politikai-társadalmi-gazdasági közeggel, amiben a szerv működött. Schellenbergnél a levéltári anyag értékelésénél a hangsúly már nem az iratok tartalmán volt, hanem az iratképzőkre vonatkozó információkon, amit manapság kontextusnak nevezünk, így téve lehetővé a nagyarányú selejtezést.<sup>13</sup>

A harmadik választóvonalhoz most értünk el, és ezt az új korszakot az elektronikus iratok problematikája jellemzi. Amint azt látni fogjuk, minden megkérdőjeleződik, új megközelítést kíván az irat fogalmától kezdve az állományvédelemig bezárólag minden. Jelen munka ezt a kihívást vázolja és ismerteti a lehetséges válaszokat.

## 2.2 Informatika, kommunikáció és elektronikus iratok

Egy bizonyos szempontból három, jól elkülönült periódust különböztethetünk meg az informatika fejlődésében, amely megfelelően leírható a hegeli dialektikával: 1) az első korszak (1943-1980) a nagy, centrális számítógépek ideje volt, amelyek nagy adatbázisokat állítottak elő (tézis); a második (1981-1995) a személyi számítógépek korszaka volt, amikor az informatika „demokratizálódott”, az adatok decentralizálódtak, és sok-sok millió PC-n osztódtak szét (antitézis); 3) a harmadik korszakot (1995-) az internet forradalma fémje-

---

<sup>13</sup> Schellenberg a levéltári információra és annak értékére vonatkozó elméletét több munkájában is kifejtette, így a *Modern Archives, Principles and Techniques* címűben (Chicago 1956), valamint a *The Appraisal of Modern Records*. In: *Bulletin of the National Archives* No. 8 (National Archives Publication No. 57-5, 1956) c. művében is.

lezte, ami összekötötte a decentralizált adatokat (szintézis). Ami ezt az utolsó periódust illeti, az két alkorszakra osztható, és a választóvonal az iPhone (2007) megjelenéséhez köthető, amikor a mobil számítógépek (okostelefonok, laptopok) aránya rohamosan emelkedni kezdett a fix gépek rovására.

A különböző periódusok és alperiódusok különböző kommunikációkat és elektronikus dokumentumokat produkáltak. Manapság az emberi kommunikáció gyökeres változást éljük át. Nemcsak egyre több az információ és egyre gyorsabb az elérési sebesség, hanem bizonyos demokratizálásról is beszélhetünk: egyre több ember lép ki a netre és hagy nyomot (magánweblapok milliói, kommentek, blogok stb.), és ezek egyre gyorsabban változnak. Tulajdonképpen az emberi beszéd és tevékenység egyre teljesebb lenyomatát találhatjuk a weben. Az instant messaging vagy akár a kommentek funkciójukban és stílusukban is (ez utóbbi nem véletlen) sokkal inkább hasonlítanak egy (telefon)beszélgetésre, mint az évszázadok alatt kialakult írott információra. Míg korábban tudomásul vettük, hogy a beszélt információ archiválhatatlan, most arra törekszünk, hogy a digitalizált verbális típusú információkat megőrizzük. Ez azonban kettős dilemmát okoz. Egyrészt nem tudunk mindent megőrizni, ezért a robbanásszerűen növekvő mennyiség miatt egyre több e-iratot kell megsemmisítenünk, és mint ahogy azt az 1.2. fejezetben láttuk, a levéltári irat értéke soha sem nulla, tehát mindig bizonytalanok leszünk, hogy nem semmisítünk-e meg értékes információt. Másrészt komoly etikai és jogi kérdést vet fel, hogy a köz- és a magán-szféra között egyre inkább elmosódik a határ, a „magánbeszélgetések” hivatalos archiválása viszont sérti az emberi jogokat. (Erre vonatkozólag később majd látunk példát.)

Az e-mail és a közösségi oldalakon történő kommunikáció is részben a szóbeli kommunikáció, a kötetlen beszéd funkcióit vette át, és főként emiatt az írásbeli szabályok, a formai jegyek eltűnnek vagy elmosódnak, és ennek, ahogy később látjuk, komoly levéltári következményei vannak. A gyakorlat nagyon laza lett, kezdve az „új helyesírástól”, a formai elemek, pl. megszólítás eltűnésén keresztül a sokszínű csatolmányokig (effektusok, kép- és hangdokumentumok). Egyre nehezebb a magán és a közösségi vagy hivatalos szféra elhatárolása. A honlapok kezdetben főként intézmények és szervezetek kommunikációs eszközei voltak, manapság a magánszféra is kiterjedten használja azokat. A közösségi oldalakkal fordított volt a helyzet, először egyének, baráti társaságok, civil csoportosulások használták, majd felfedezték a közintézmények és közszerplők, pl. a politikusok. Ez utóbbiak gyakran az Facebookot használják politikai akciójuk vagy bejelentéseik elsődleges eszközeiként. A hivatalokban a munkatársak mail boxaiban keverednek a ma-

gán- és a hivatalos mailek, nem kevés fejtörést okozva az e-levelezést archiválni akaró records managereknek vagy levéltárosoknak.

Ami az irattípusokat illeti, a kezdeti nagy, központi számítógépek korszakában szinte kizárólag egyszerű szerkezetű, nagy adatbázisokat hoztak létre. A PC-k korszakában a szövegszerkesztők és a táblázatkezelők által képzett iratok dominálták a digitális világot. Az internet korában az elektronikus üzenet, különösen az e-mail az uralkodó, de ugyanakkor sokféle és sokszínű, különböző típusú iratot láthatunk köszönhetően az újfajta kommunikációs eszközöknek, a közösségi hálóknak, amelyek kifejllesztették a saját stílusukat és formájukat, és amelyek jelentősen eltérnek a tradicionális iratoktól (levelek, táviratok stb.). Végül de nem utolsósorban a web világában az e-iratok egyre nagyobb része átmeneti vagy dinamikus (gyorsan, állandóan változó) irat. A magániratoknál (ilyen pl. egy személy weblapja) ez magától értetődő, de hasonló a tendencia a hivatalosaknál is.<sup>14</sup> Egyrészt, mert könnyű a létrehozás és a publikálás, ezért pl. egy hivatalos intézmény naprakész információkkal látja el az ügyfeleit, vagy az ügyintézés (pl. regisztrálás, vásárlás) egyre inkább a webre tereli. Másrészt civil vagy akár hivatalos szervezetek egyre gyakrabban vonnak be egyre több embert az iratok, adatok létrehozásába, ennek egyik legjobb példája a Wikipedia, de ilyenek a genealógiai adatgyűjtemények egyesítésére tett kísérletek vagy a később (12. fejezet) részletesebben ismertetett BBC Domesday-projektje. (A rendkívül interaktív és személyre szóló web szolgáltatásokat nevezik Web 2.0-nak.)

### 2.3 Az emberi emlékezet digitalizálása és adatrobbanás

Az utolsó 50 évben az informatika fokozatosan meghódította az emberi emlékezetet. Kezdetben a nagy nyilvántartásokat és katalógusokat „gépesítették”, aztán a szöveges és a táblázatos anyagok következtek, majd a rajzok. A következő etap a képek digitalizálása volt (mind az álló-, mind a mozgóképeké), majd a hangdokumentumok jöttek, és a nemrég megjelent 3D-s nyomtatókkal elérkeztünk a fizikai objektumok digitalizálásához. Manapság már elmondhatjuk, hogy az informatika képes létrehozni vagy reprodukálni az emberi emlékezet tárgyainak többségét.

Az eredmény az e-adatok robbanásszerű növekedése, ami egy önmagát erősítő folyamat:

---

<sup>14</sup> Anne J. Gillilan-Swetland: Testing our Truths: Delineating the Parameters of the Authentic Archival Electronic Record. In: *American Archivist* vol. 65 Fall/Winter 2002, 207-208. old.

- Mivel majdnem minden emberi kommunikáció (szöveg, kép, hang) digitalizálható, ezért egyrészt egyre több tevékenység helyeződik át a digitális világba (pl. levélírás, beszélgetés), másrészt gyorsan növekszik az „adatgyártók” száma (mindenki, aki e-mail-fiókot vagy facebook profilt, fotóalbumot stb. csinál, már e pusztán tényével is adatgyártóvá válik).
- Az adatokat nagyon könnyű létrehozni még manuálisan is (digitális szövegírás, szerkesztés, hangrögzítés, fotózás stb.), de különösen automatikusan. Az adatok egyre nagyobb hányada generálódik valamilyen informatikai eszközzel, kezdve a műholdak vagy meteorológiai műszerek mérési eredményeitől az e-mailek küldésekor automatikusan rögzített metaadatokig (pl. a küldőre és a címzettre vonatkozó adatok) bezárólag.

Jóllehet a papír nélküli iroda időszaka még nem érkezett el, de az idő közeleg. A személyi számítógépek idején az elektronikus (szövegszerkesztett) iratok többségét még kinyomtatták, ami ma már nem jellemző, jelenleg egy átmeneti korban élünk. A kormányzatok, főként takarékosági és praktikus okok miatt gyakran arra bátorítják a szervezetet, hogy a papír formájában kapott iratokat is digitalizálják, és így integrálják azokat az elektronikus rendszerükbe. Az állampolgárok pedig azzal szembesülnek nap mint nap, hogy egyre több szolgáltatás – államiak is – már csak interneten keresztül érhető el digitális formában.

#### 2.4 Az elektronikus iratok tartós megőrzésének nehézségei

Az írásbeliség története azt mutatja, hogy a fejlődés során a hordozók (kötőlap, agyagtábla, papirusz, pergamen, papír [könyv], mikrofilm, mágnes- és optikai lemez), bár kisebb-nagyobb eltérésekkel, de egyre sűrűbb információt tartalmaztak és egyre sérülékenyebbek lettek, már csak a sűrítés miatt is. A sűrítés a mikrofilmnél már elért egy olyan pontot, amikor az információ csak segédeszközzel volt olvasható. Az analógról a digitális írásbeliségre, információrögzítésre való áttérésnek – ami a 20. század második felében kezdődött – két fő hajtóereje volt: a további információsűrítés és a jobb információkezelés (-rendezés, -visszakeresés). Viszont a digitalizált információ olvasásánál a segédeszköz (hardver, szoftver) már nem csak a nagyításhoz kell, hanem a megértéshez is, mert az önmagában már érthetetlen az ember számára, kódolva van. Ebből kifolyólag már nem elegendő a fizikai megőrzés, a „dekódoló szótárt” is meg kell őrizni, ami nem feltétlenül az eredeti hardver vagy szoftver. Az iratoknak az analóg világban gyakorolt hosszú távú hiteles megőrzése a digitális világban lehetetlen, és ennek két fő oka van: technikai és levéltártani. A technikai okok jól ismertek, és az előbbiekben már utaltam rá: 1) az e-iratok képzéséhez,

működtetéséhez (olvasásához) és megőrzéséhez technikai eszközrendszer (hardver, szoftver és e-környezet) szükséges, ami értelemszerűen determinálja az e-iratokat; 2) mivel az eszközrendszer az idők folyamán gyorsan változik, ezért az iratokat lehetetlen megőrizni az eredeti állapotukban.

Mindenki, aki hosszabb ideje használ számítógépet, tapasztalhatta, hogy a régi lemezek, fájlok az újabb gépen vagy újabb szoftverekkel nem olvashatók-használhatók, vagy többé-kevésbé hibásan működnek, mert már nem felelnek meg az aktuális szabványoknak. Ahogy nő az időintervallum, úgy csökken az esély a használatra. 20 évvel ezelőtt általános volt a floppy lemezek használata, a mostani tizenévesek legfeljebb hallottak róla. Manapság egyre több számítógép készül CD-DVD-meghajtó nélkül, nem beszélve az új okostelefonokról (ezek is számítógépek!), ahol fizikailag nincs is hely az ilyen médium használatára. A DOS-os vagy korai Word-del szövegszerkesztett fájlokat legfeljebb szakemberek tudják életre kelteni. A hardver-, illetve szoftverrelavulás miatt 5-10 év után veszélybe kerülnek az e-iratok.

A szabványok hiánya vagy a túl sok szabvány, ami lényegében ugyanazt eredményezi, jelen időben is fennáll, nemcsak az idő múlásával. Ha valakinek nem áll a rendelkezésére a megfelelő szoftver, akkor nem tudja olvasni az adott szoftvertől függő fájlokat. A szabványosság, mint működési követelmény egyébként minden környezetben fennáll, még az analóg világban is, de ott természetesnek tartottuk – igazából észre sem vettük –, hogy minden emberi szem „szabványos”, tehát gond nélkül tudtuk olvasni az analóg iratokat. (Hogy értettük-e, az más kérdés.) Az e-iratoknál döbrentünk rá, hogy a „szabványtalanságnak” milyen nagy a kockázata, milyen súlyos a következménye – sokkal nagyobb, mint ugyanez a probléma más technikai környezetben (pl. autókban).

Az e-iratok hosszú távú hiteles megőrzésének azonban nemcsak technikai, hanem levéltártani feltételei is vannak. Ezek már kevésbé ismertek, pedig legalább olyan fontosak, mint a technikaiak, és legalább olyan sok problémát okoznak, mert az informatika jelentősen átfogalmazta az archivisztika korábbi fogalmait és szabályait. Minden lényeges levéltártani változás eredője négy dologra vezethető vissza: 1) az e-iratnak nincs fizikai aspektusa; 2) az e-iratnak megváltozott a szerkezete és erősen függ a technikától; 3) az e-iratot könnyű létrehozni, változtatni, sokszorosítani, kombinálni és terjeszteni; 4) az előző okok miatt robbanásszerűen növekszik az e-iratok mennyisége.

Az e-iratok hosszú távú megőrzésének célja, hogy megőrizzük az iratok valós és potenciális értékét, hogy ezáltal a levéltárak be tudják tölteni a funkciójukat, ahogy azt az 1.2.

fejezetben láttuk. Nem öncélú levéltári vagy könyvtári követelmény ez, hanem a társadalom, sőt az emberi civilizáció elemi érdeke, és reményeim szerint ezt be tudom bizonyítani a könyvben. A következő fejezetekben ezeket a kérdéseket fejtem ki részletesen.

### 3 A levéltári irat és alkotó elemei

Az irat nem azonos az információval vagy az adattal, annál komplexebb fogalom, egy adat vagy információ önmagában még nem irat, sok követelménynek kell megfelelnie, hogy azzá váljon. Sokféle irat van, a témánk szempontjából számunkra a legfontosabb a levéltári irat fogalma és felépítése. A levéltári szakirodalom nagyjából egységes elvek mentén határozza meg az irat fogalmát. Idézzük a Nemzetközi Levéltári Tanács levéltártani terminológiai szótárának a fondra vonatkozó definícióját, amely – mivel magában foglalja –, érvényes a levéltári iratra is: a fond „egy természetes vagy jogi személy által készített vagy kapott iratok együttese, amelyeket [a létrehozó] bizonyítékként vagy információs céllal őriz akár jogi okokból, akár tevékenységének gyakorlása céljából.”<sup>15</sup> Ebből a definícióból viszont egy fontos dolog hiányzik: az iratoknak az irattári rendszerbe regisztrálnak (iktatottak) kell lenniük.

A levéltári irat több, egymással kapcsolatban álló alkotóelemből áll: hordozóból, formából, tartalomból, azonosítóból, struktúrából és kontextusból.<sup>16</sup>

#### 3.1 Hordozó

Az analóg világban hordozónak tekintjük azt a fizikai tárgyat, ami az információt hordozza, a megjelenítését lehetővé teszi. Papíriratok esetén ez magát a papíryanagot, illetve a tintát vagy nyomdafestéket jelenti, és a „hordott” információt közvetlenül, emberi szemmel tudjuk olvasni. Mikrofilmek esetében már egy kicsit bonyolultabb volt a helyzet, mert a kényelmes olvasásához szükséges egy gép, persze ez végső soron lehet egy nagyító is, de a hordozó fogalma alatt itt is a fizikai objektumot (filmtekercset vagy filmlapot) értünk.

Az e-világban mindez gyökeresen megváltozott, főként mert az e-iratnak nincs fizikai aspektusa. Ez a felismerés igazán az internet elterjedésével vált világossá sokak számára: az e-iratot őrizhetik bárhol a világon, lehet bármilyen fizikai hordozón (mervlemezzen, CD-

---

<sup>15</sup> <http://www.ciscra.org/mat/termdb/term/4319/5205>, a definíció a MOREQ2-től lett kölcsönözve.

<sup>16</sup> Vannak más felosztások is. Pl. az InterPARES-projekt az e-iratoknak – főként a diplomatikára alapozva – négy alapvető elemét különböztet meg: dokumentációs forma, annotációk, médium és kontextus. Lásd: *Requirements for Assessing and Maintaining the Authenticity of Electronic Records, Appendix 2*. Elérhető: [http://www.interpares.org/display\\_file.cfm?doc=ip1\\_template\\_for\\_analysis.pdf](http://www.interpares.org/display_file.cfm?doc=ip1_template_for_analysis.pdf)

n, pendrive-on stb.), bármilyen kiosztásban (pl. szétszórva több fájlban). Ráadásul a digitális információt képviselő bitek ugyan jelen vannak egy merev- vagy optikai lemezen, esetleg a számítógép memóriájában, ez azonban önmagában nem elegendő az információ megjelenítéséhez, azaz „hordozásához”, ehhez még egyéb technikai eszközrendszer (hardver, szoftver és egyéb e-környezet) is szükséges.<sup>17</sup> A digitális világban az e-irat szélesebb értelemben vett hordozója tehát ez az eszközrendszer, ami részben fizikai, részben virtuális. Hogy a dolgok még bonyolultabbak legyenek, ahogy láttuk a 2.4. fejezetben, a technikai fejlődéssel ez a hordozó állandóan változik, az e-irat többi alkotó eleme pedig tőle elválasztható (átmásolható teljes azonosságban más hordozóra, konvertálható más szabvány szerint). Az analóg világgal ellentétben a hordozó azonossága már nem feltétele az irat megjelenítésének és ebből következőleg a megőrzésének, sőt, mint látni fogjuk, hosszú időtávban az egyenesen lehetetlen. A hordozó kitüntetett szerepe az e-világban ezért jelentősen csökkent.

### 3.2 Forma

Az irat formája tulajdonképpen a kinézetét jelenti: írott dokumentumnál az, hogy milyenek a színei, a méretei, milyen betűk vagy ábrák vannak benne, van-e fejléce stb. Az analóg világban ezek még kiegészülnek fizikai formai jegyekkel: pl. tekercs vagy könyv alakú az irat. (A hang- vagy képdokumentumoknál a forma természetesen mást jelent, de erre most szükségtelen kitérni.) Az analóg iratoknál a forma elválaszthatatlan a hordozótól és automatikusan észleljük, értjük. Az e-világban – maradvány az írott dokumentumoknál – ugyanezeket már a hardver és a szoftver (pl. szövegszerkesztő vagy adatbáziskezelő) kiválasztásával erősen befolyásoljuk, illetve a dokumentum szerkesztésekor határozzuk meg, pl. hogy milyen betűtípust alkalmazzunk, legyen-e fejléc vagy sem. Olvasáskor a hardver és a szoftver pontos technikai specifikációk (metaadatok) alapján végzi el a megjelenítést, és ha nem teljes a kompatibilitás, pl. az olvasó szövegszerkesztője nem rendelkezik a megfelelő karakterkészlettel, akkor megváltozik a forma.

A forma lehet ugyanabba a fájlba integrálva, amiben a tartalom van, de akár el is válhat tőle, pl. egy külön fájl (stylesheet) tartalmazza azt, illetve egy tartalomhoz akár több forma (stylesheet) is rendelhető, ez pl. az XML-dokumentumoknál előfordul. Mivel a forma elválik a tartalomtól, de erősen függ a hardver és szoftver egyes elemeitől, ami viszont változik alkalmazónként és az idők folyamán, az e-világban valamelyest csökken a forma szerepe,

---

<sup>17</sup> Bár a szakirodalomban hordozó alatt többnyire magát a médiát értik, így pl. az előbb említett ICA terminológia vagy az OAIS is, de véleményem szerint az e-világban a hordozó fogalma is kitágult.

pontosabban a hosszú távú megőrzésnél a levéltárosok kénytelenek bizonyos kompromisszumokat kötni, és nem ragaszkodnak a forma teljes megőrzéséhez. Ez történik, amikor konvertálnak, vagy amikor csak az irat lényegének (essence) megőrzéséhez ragaszkodnak, ami gyakran bizonyos nem lényeges formai elemek feláldozását jelenti. A forma fontossága irattípusként is változhat, pl. egy grafikonnál általában nagyon jelentős, de egy adatbázisnál már kevésbé fontos, hogy milyen típusú betűkkel jelenítjük meg a találatokat.

### 3.3 Tartalom

A tartalom az irat tárgyára vonatkozó immanens információ. Funkciója és jelentősége ugyanaz az analóg, mint az e-világban, lényegében azért, mert mindkét világban demateriális, tehát az e-világ szubsztanciájával egybeesik. Meg kell jegyezni, hogy mind a formának, mind az alább ismertetendő kontextusnak komoly tartalmi vonatkozása is van, de mivel egyik sem immanens, és nem csak az irat tárgyára vonatkoznak, ezért külön tárgyalom őket.

### 3.4 Azonosító

Az azonosítónak kettős szerepe van: 1) az iratot egyedileg azonosítja, lehetőleg úgy, hogy azt a jövőben ne kelljen megváltoztatni; 2) az egyedi azonosítással lehetővé teszi az iratrendszerben elfoglalt helyének pontos kijelölését, ezzel nemcsak azt jelezve, hogy mely irategyüttesekbe tartozik, hanem azt is, hogy mely iratok tartoznak össze. Ezért látják el a levéltárak, kéziratárak egyedi jelzettel az irataikat. Az egyedi és fix azonosító fontos követelmény az irat identitásának (egyediségének) és hitelességének megőrzéséhez. Az egyedi azonosítót egy rugalmas rendszer alapján érdemes létrehozni, ami képes utólag befogadni új iratokat, külső azonosítókat, új típusú dokumentumokat.

Az azonosítók jelentősége nagymértékben megnőtt az e-világban. Egyrészt napi gyakorlattá vált különböző intézmények adatbázisainak, digitalizált iratainak integrációja pl. egy levéltári portálon, és a nem megfelelő, nem egyedi azonosítók komoly zavart okozhatnak. Ezért vannak nemzetközi törekvések az egyedi azonosítók szabályainak kialakítására.<sup>18</sup> Másrészt, amint azt a következő fejezetben látni fogjuk, az e-iratok struktúrája jellemzően sokkal bonyolultabb, mint az analógoké, ezért számtalan olyan új, a felhasználó számára gyakran láthatatlan azonosítót is alkalmaznak, amelyek az analóg világban ismer-

---

<sup>18</sup> Már a levéltári leírásra kidolgozott nemzetközi szabvány, az 1994-ben megjelent ISAD(g) is hangsúlyozta az egyedi azonosító jelentőségét. Az informatikai világban pedig több szabvány létezik az egyedi azonosítókra, főként azért, hogyha a digitális objektum helye megváltozik, akkor az új helyen is lehessen azonosítani. Ilyen pl. a Universal Resource Names (URN) vagy a Digital Object Identifiers (DOI) vagy a Persistent Uniform Resource Locators (PURL).

retlenek voltak. Egy nagyobb adatbázis adatait akár száznál is több adattáblába rendezve tárolhatják, a kezelő szoftver pedig csak megfelelő azonosítók esetén tud „rendet tartani”. Ezeknek a „láthatatlan” azonosítóknak a megőrzése nélkülözhetetlen egy későbbi konvertálásnál. De vannak egyéb, olyan technikai azonosítók is, amelyek az irat (a fájl) e-környezetét írják le, pl. a formátumát, amelyek egy későbbi konvertálásnál nélkülözhetetlenek. (Ezek az azonosítók, más szempontból nézve, kontextusnak számítanak.)

Az azonosítókat többnyire metaadatként kezeljük, és a tartalommal és/vagy a formával – akár a fájlnevbe – integrálva lehetnek egy fájlban, de lehetnek külön táblába is kigyűjtve.

### 3.5 Struktúra

Az iratok, mivel tartalmi, formai, logikai és informatikai szempontból is összetettek, ezért szerkezettel bírnak, amit többféleképpen lehet osztályozni, pl. oklevéltani, levéltártani vagy információelméleti szempontból. Én a levéltártani struktúrát állítom a középpontba, mert ez hordozza azt a koherenciát, ami az iratok integritásának hosszú távú megőrzésénél elengedhetetlen. Egy iratnak van belső és külső szerkezete, és az iratok külső szerkezeteiből állnak össze az irategyüttesek.

#### 3.5.1 *Belső struktúra*

Az információk, adatok irattá strukturálódnak, és ezt a belső szerkezetet legalaposabban a diplomatika tárta fel. A régi oklevelek szerkezetéből sok átöröklődött a későbbi (jelenkori) iratokra is, ami már az irattan és részben a levéltártan felségterülete. Egy hagyományos levélre elég csak rápillantanunk és látjuk-értjük a címzettet, a tárgyat, a dátumot, az aláíró – azaz a belső struktúrát –, egy e-levél esetében, amikor kitöltjük pl. a címzett- vagy tárgymezőt, a szoftver automatikusan generál olyan metaadatokat, amelyek segítségével az olvasó számítógépe ugyanilyen elrendezést rekonstruál. A papírvilágban címek és alcímek kiemelésével érzékeltetni tudjuk egy irat vagy egy irategyüttes struktúráját, egy weblap esetén ezt külön kell jelezni a számítógépnek, pl. címkékkel (tag-ekkel).

E-iratok esetén mindezekhez olyan technikai adatok járulnak, amelyek az analóg világban nem léteztek vagy nem tekintették őket lényegesnek. Ilyen pl egy e-mail esetén az adatátvitelre vonatkozó információk: az elküldés és a kézhezvétel időpontja, az e-mail útvonala a küldőtől a fogadóig stb.

Az e-világban az iratok belső szerkezete erősen veszélyeztetett, különösen konvertálás esetén. Ekkor eltűnhetnek olyan metaadatokat, amelyek nélkül az irat értéke erősen degradálódik.

### 3.5.2 Külső struktúra

A levéltári iratok kapcsolódnak más iratokhoz: tartalom, eljárás, funkció vagy forma szerint csoportosulnak, így az egy ügyhöz tartozók ügyiratot alkotnak, az ügyiratok csoportja tételként jelentkezik az irattári tervben, a tételek sorozatot formálnak, több sorozat alkot egy állagot, az állagok csoportja pedig egy fondot. Ez alapján megkülönböztethetünk iratszintet, ügyiratszintet, tételszintet, sorozatszintet, állagszintet és fondszintet. Az egységek közötti kapcsolatok általában megfelelnek az eredeti irattári rendnek, ez pedig – többé-kevésbé – tükrözi az iratképző működését.

A külső struktúra azonban nem áll meg fondszinten. Egy levéltár általában sok helyről (iratképzőtől) gyűjti az iratait, azaz sok fondot őriz, ezeket pedig egy nagy struktúrába kell integrálnia, különben nem tudja kiépíteni a nyilvántartó és az információ-visszakereső rendszerét. Ezt a fond fölötti szerkezetet nevezhetjük makro- vagy külső struktúrának. Az egyes iratok azonban közvetlenül is kapcsolódhatnak más csoportba tartozó iratokhoz. Erre a legjobb példa egy weblap, ami linkek révén van összekötve más weblapokkal.

A külső struktúrában a kapcsolatok iránya kétféle lehet: vertikális, illetve horizontális. A vertikális jelző az összefüggés irányára utal, mert ha lerajzoljuk egy levéltár irategyüttesei (állag, fond, fondcsoport stb.) közötti kapcsolatokat, akkor a hierarchikus kapcsolatokat függőleges vonalakkal tudjuk érzékeltetni. A horizontális jelző „egyenrangú” kapcsolatokra utal, pl. arra, hogy mely sorozatok alkotnak egy állagot, vagy hogy egy e-mail milyen korábbi e-levelekre válasz, illetve milyen papíriratokkal kapcsolatos.

Az iratok információs értékének megőrzéséhez elengedhetetlen a külső struktúra megőrzése. Ha pl. egy iratsorozat darabjait szétszórjuk és nem jelezzük az összetartozásukat, akkor a fizikai sértetlenségük ellenére az értékük (mindhárom) jelentősen csökken, mert nem köthetők a keletkeztetésének okához, az ügyhöz, a szervezethez vagy személyhez, amely létrehozta őket. Ezért fontos a struktúrának a feltárása, dokumentálása és megőrzése, az analóg világban ezt szolgálták a különböző nyilvántartások, leírások és segédletek.

Az e-világban bonyolultabb a helyzet, mert az e-iratoknak jellemzően sokkal több külső kapcsolódásuk van: a weblapok tele vannak linkekkel, az e-mailek csatolmányokkal. Már az az egyszerű tény, hogy a háromdimenziós világ az informatikában kétdimenziósra szűkül, komoly hátrányt jelent, amit csak strukturális metaadatokkal tudunk ellensúlyozni. A hagyományos levéltári iratok kutatásánál a dobozokat, dossziékat – azaz a külső struktúrát – azonnal felismerjük, hiszen elkülönülnek a többi doboztól és dossziétól, a bennük

lévő lapokat forgathatjuk, csoportosíthatjuk az általunk kívánt logikai rendbe. (Aztán a kutatás végén, természetesen, helyre kell állítani a levéltári rendet.) Ha ugyanezeket az iratokat digitalizáljuk – ilyenkor egy-egy oldal egy-egy képfájlba kerül –, akkor azok a képernyőn szigorúan szekvenciális rendben sorakoznak és sokkal nehezebb őket kezelni. Ezt ellensúlyozandó rendelkez adatbázist, azaz metaadatok együttesét a szkennelt képek mellé, amelyek mutatják az összetartozást vagy az elkülönülést, és melyek segítségével lehet hasonlóképpen – sőt jobban – kezelni, böngészni, keresni az iratokat.

### 3.6 Kontextus

A levéltári szakirodalom kontextusnak nevezi az iratképzőre vonatkozó adatokat, információkat (a szerv vagy személy [élet]története, tevékenysége, kompetenciája, funkciói, szervezeti felépítése [szerv esetén], kapcsolata más szervekkel vagy személyekkel stb.), valamint az iratképzővel, illetve az irattal-irategyüttessel kapcsolatos jogi, politikai, társadalmi vonatkozásokat.<sup>19</sup> (Pl. ha egy iratsorozat egy jelentős társadalmi eseményről szól, akkor röviden meg kell világítani az esemény hátterét.) Ezek az összefüggések gyakran csak impliciten vannak az iratokban, ahonnan a levéltárosnak kell kibontani azokat, sőt sokszor kiegészítésre szorulnak más forrásokból. A levéltári kontextus jellemzően az iratok leírásakor kerül rögzítésre metaadatként. (Erről részletesen a 4. fejezetben lesz szó.)

Az e-irat egyben digitális objektum is, ezért fontos technikai attribútumai vannak – ilyenek pl. a fájlformátumra vonatkozó adatok –, amelyek mind a megőrzésnél, mind a későbbi használatnál nélkülözhetetlenek. A technikai kontextusnak – ami az informatikai környezetet írja le – különösen részletesnek kell lenni akkor, ha a levéltár alkalmazni kívánja az emulálást (lásd a 9.2. fejezetet!), ekkor mind a hardver, mind a szoftverkönyezetet egészen részletesen kell dokumentálni. A technikai követelmények miatt a kontextusnak, mint iratot alkotó elemnek a jelentősége megnőtt az e-világban.

Mindemellett vannak még olyan kontextusai az iratoknak – mind a hagyományosoknak, mind az e-környezetben születetteknek –, amelyek nem tartoznak a tárgyunkhoz, de feltétlenül említést érdemelnek. Az iratok közötti (főleg tartalmi) összefüggésekre gondolok,

---

<sup>19</sup> Jó összefoglalót ad a levéltári kontextusról az *ISAD(G): Norme générale et internationale de description archivistique* (Conseil international des archives, Ottawa 2000 – a szabvány korábbi verziójának magyar fordítását lásd a Levéltári Szemle 1995/2. számában), részletesebben kifejti a kontextus egy részét az *ISAAR (CPF): Norme Internationale sur les notices d'autorité utiles pour les Archives relatives aux collectivités, aux personnes ou aux familles* (Conseil international des archives 2004, a magyar fordítását lásd a Levéltári Szemle 2005/3. számában); elérhetők: <http://www.ica.org/10241/normes/liste-des-normes.html>. Az InterPARES csoport más szempontból ötféle kontextust különböztet meg: jogi-adminisztratív, provenienciára vonatkozó, eljárási, dokumentációs, technológiai. Lásd Luciana Duranti, Kenneth Thibodeau: *Interpares 2: Experimental, Interactive and Dynamic Records, Appendix 2*, 18. old. Elérhető: [http://www.interpares.org/ip2/display\\_file.cfm?doc=ip2\\_book\\_appendix\\_02.pdf](http://www.interpares.org/ip2/display_file.cfm?doc=ip2_book_appendix_02.pdf)

amiket a pontos elhatárolás kedvéért nevezhetünk kutatói kontextusoknak. Ezek általában rejtett kapcsolatok, amelyek kibontása elsődlegesen nem a levéltáros, hanem a kutató, az anyagot használó feladata. A kutatómunka lényege tulajdonképpen ezeknek az összefüggéseknek a felderítése. Minél rendezettebb, jól strukturáltabb, pontosabb és részletesebb a leírt levéltári és technikai kontextus, annál nagyobb esély nyílik a kutatói kontextusok feltárására.

### 3.7 Az elektronikus iratok alkotó elemeinek szétválása

Az analóg világban többnyire konszenzus volt, hogy a tradicionális iratot a hordozó (pergamen, papír, tinta, nyomdafesték stb.), a forma és a tartalom együttese alkotja, ezek szétválaszthatatlanok, tehát a megőrzés mindhárom alkotóelem együttes megőrzését jelenti. A hordozó, a forma és a tartalom összetartozásával kapcsolatos felfogást csak csekély mértékben kezdte ki az analóg másolatok terjedése, így az elektrosztatikus másolatok térhódítása, illetve a mikromásolatok (mikrofilmek) alkalmazása. A mikrofilm a szerveknél keletkezett szöveges iratok információjának 99%-át rögzítette – de pl. a papír vízjelét vagy a nem fekete-fehér színeket nem –, nagyon helytakarékos és könnyen használható volt, mégsem tudott elterjedni az ún. helyettesítő mikrofilmezés, amikor a filmre vett iratokat a filmezés után kiselejtezték, ellentétben a biztonsági mikrofilmezéssel, amikor mind az eredetit, mind a filmet megőrizték. Bár sokak szerint ennek fő oka a jogi hiányosságban volt keresendő – az egyedül megmaradt mikrofilmek jogi bizonyító ereje több országban nem volt egyértelmű –, itt inkább a fentebb vázolt társadalmi szemléletről volt szó, amit a jog is tükrözött: ha a hordozó megváltozik, az sérti a hitelességet.

A digitális világban mindez gyökeresen megváltozott. Amint az előző fejezetekben láttuk, az irat dematerializálódott, az alkotó elemeinek fogalmi átalakultak, kiegészültek technikai attribútumokkal, egymástól elváltak (vagy elválhattak) és le- vagy felértékelődtek. Mindazon elemek, amelyek rendet visznek ebbe a komplikált rendszerbe, mint az azonosító, a struktúra és a kontextus, felértékelődtek, mert az irat identitásához, integritásához és hitelességéhez nélkülözhetetlenek. Ezek a változások az iratokat még bonyolultabbá és sérülékenyebbé tették – gondoljunk pl. arra, hogy mi történik, ha a külön szereplő elemeket fájl sérülés miatt nem sikerül „összehozni” –, ezért is olyan nehéz az e-iratok hosszú távú megőrzése.

### 3.8 A levéltári iratok egyedi jellege

A levéltári iratok felépítése nemcsak bonyolult, hanem sajátos is, és lényegesen különbözik például a könyvtári vagy a múzeumokban őrzött iratoktól. Sajátossá teszi őket egyrészt a kapcsolatrendszer, különösen a külső, hierarchikus struktúra, másrészt a levéltári kontextusok léte. Mindezek összességben rendkívül komplikálttá teszik a levéltári anyag feldolgozását, sőt a kutatást is. Nem öncélúan alakult ki ez a rendszer. Amint fentebb láttuk a levéltári szemlélet sarokköve a fond. A fond

- jellemzően komplex adatokat és információt tartalmaz magáról az iratképzőről és környezetéről, tükrözve az érintett személyek és dolgok közötti kapcsolatokat;
- iratai eljárásokban képződnek, azaz időfolyamban; az eltelt idő alatt az iratképző és a környezete változik, amely változást az iratok tükrözik.

Mind a komplexitás, mind a folyamat csak kapcsolatokkal (struktúrával) és kontextussal képezhető le és írható le. A levéltáros az iratanyagot fondtól lefelé, azaz az általánostól az egyediig elemzi, így írja le azokat, és mindig szem előtt tartja irategyüttesek összetartozását, azaz intellektuális egységét. A levéltárakban alkalmazott tipikus kutatási módszerek is alkalmazkodtak a fent leírt rendszerhez. A levéltári információ tartalmi részét általában a vertikális kapcsolatok révén éri el a kutató: a keresett iratot lépésről lépésre, szintről szintre, felülről lefelé, egyre szűkebb körben lokalizálva közelíti meg. Ha egy kutató a levéltáros segítségét kéri egy adott témára vonatkozó irat megtalálásához, a levéltáros általában azt keresi, hogy a témában vajon melyik szerv volt kompetens, aztán a szervezen belül melyik osztály stb. A levéltári terminológia ezt az eljárást néha hierarchikus, néha közvetett (indirekt) megközelítésnek nevezi, mindenesetre szöges ellentétben áll a tipikus könyvtári információ-megközelítéssel, amit közvetlennek (direktnek) nevezhetünk. Ha egy (elektronikus) könyvtári katalógust használunk, akkor vagy a szerző neve, vagy a mű címe, vagy a besorolása alapján keresünk, és azonnal elérjük az információt (már ha sikeres a keresésünk).

Mindez persze nem zárja ki a levéltárakban sem a közvetlen információ-megközelítést, azaz az indexek használatát, sőt ezek jelentősége az e-korban növekedett, mert könnyebbé vált a részletes, akár darabszintű indexálás és az erre alapozott keresés. Az azonban nem változott, hogy a tárgyi indexálás rendkívül szubjektív (lásd a 2.1. fejezetet), és egy idő után biztosan elavul, ezért a jövőben sem nélkülözhető az információ indirekt megközelíthetősége. Ezért olyan fontos a levéltári adatbázisokban a hierarchikus böngészés-funkció.

## 4 A metaadatok

A metaadatok a közkeletű definíció szerint adatok az adatokról (iratokról). Metaadatokat az analóg világ is ismert, csak nem így hívta őket, hanem pl. könyvtári katalógusnak vagy levéltári segédletnek. Az e-világban azonban lényegesen megnőtt a szerepük, mert az iratalkotó elemek többnyire metaadatként jelennek meg, így pl. (technikai) metaadattal hivatkozhatunk a hordozóra, metaadat az azonosító, metaadatokkal írjuk le a struktúrárt és természetesen a kontextust. Az elektronikusan született iratokhoz ugyanúgy hozzátartoznak, mint például a digitalizált iratképekhez. Az utóbbiaknál, valamint a hangfájlok esetén különösen fontosak, mert a szövegfájlokkal ellentétben ezekben nem lehet szavakra, kifejezésekre keresni, tehát a nyilvántartásnak vagy a keresésnek egyetlen eszköze a metaadat. Fontosak a metaadatok a nagy tudományos adatbázisoknál is, mert általuk írják le, hogy pl. hogy gyűjtötték, kalibrálták a műszereket, a folyamat alatt milyen paramétereket alkalmaztak stb., tehát ezek a jövő tudományos kutatásainak zálogai.<sup>20</sup>

Fontos jellemzőjük a metaadatoknak, hogy bár emberi szemmel olvashatóvá tehetők, a fő funkciójuk az, hogy géppel is feldolgozhatók legyenek (kereshetők, rendezhetők stb.).<sup>21</sup>

A metaadatokat többféleképpen kategorizálják, a szakirodalomban a legelterjedtebb a Making of America II Testbed Project által még a 1990-es években felállított hármas csoport:<sup>22</sup>

- Leírók: az iratok azonosításához és az adatfeltáráshoz szükséges adatok (ez utóbbiak az iratok tartalmára vonatkoznak, azaz segédlet vagy adatbázis jellegű adatok).
- Adminisztratívak: a digitális anyag használatához szükséges adatok (például technikai specifikációk vagy a készítés körülményeire, illetve kutatási korlátozásokra vonatkozó adatok).
- Strukturálisak: a digitális objektumok komponensei közötti kapcsolódásokra vonatkozó adatok (értendő ez alatt egyaránt a levéltári vagy a technikai kapcsolódás).

---

<sup>20</sup> Michael Day: *Instalment on "Metadata"* (DCC Digital Curation Manual 2005) 10.old. Elérhető: <http://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/metadata/metadata.pdf>

<sup>21</sup> Michael Day i.m. 12. old

<sup>22</sup> Lásd pl. Bernard J. Hurley & John Price-Wilkin & Merrilee Proffitt & Howard Besser: *The Making of America II Testbed Project: A Digital Library Model* (1999). Elérhető: <http://www.clir.org/pubs/reports/pub87/pub87.pdf>

De a metaadatokat lehet más, pl. az e-iratok életciklusának szempontjából is csoportosítani.<sup>23</sup>

- Irattáriak: ezek a szerveknél az iratkezelő rendszer működése során az elektronikus iratokkal együtt, általában automatikusan generálódnak. Tartalmazhatják a szerző nevét, a keletkezés dátumát, az irat címét, minősítési fokát, az irathoz rendelt kulcsszót stb.
- Levéltáriak: ezeket az irat lezárása után adják vagy a szervnél vagy a levéltárban beszállítás (új e-iratok fogadása) után. Tartalmazhatják az utolsó iratellenőrzés dátumát, az iratképző szerv nevét stb.
- Technikaiak: az elektronikus iratok technikai megértéséhez és kezeléséhez szükséges adatok, amelyek keletkezhetnek akár a szerveknél, akár a levéltárban. Kívánatos, hogy a szervnél automatikusan generálódjanak, levéltárban pedig csak akkor, ha ott az iratok konvertálására (migrálására) kerül sor.

A metaadatok lehetnek az iratfájlba integrálva (mint pl. a PDF/A formátumú fájlok esetén), lehetnek „összecsomagolva” az iratfájllal, vagy lehetnek külön táblába kigyűjtve. (Általában előnyösebbnek tartják, ha együtt vannak, mert így kisebb a hibalehetőség.) Ha külön vannak kigyűjtve, akkor a metaadatok együttesét adatbázisnak is tekinthetjük, nemcsak informatikai, hanem levéltártani szempontból is, ezért érdemes azokat levéltártani elveknek megfelelően felépíteni. Ilyenek pl. az iratok tartalmát, kontextusát leíró kutatói adatbázisok. A technikai metaadatok nélkülözhetetlenek a karbantartásnál és a migrálásnál (9.1. fejezet), figyelni kell, hogy milyen kifutó, elavuló formátumú fájlok hol vannak.

A 3. fejezetben láthattuk, hogy az azonosítók, a formát vagy a struktúrát leíró adatok és a kontextus az e-világban mind metaadatként jelentkeznek, tehát részei az e-iratoknak, szervezett rendszerré teszik azokat, nélkülük csak digitális objektumok használhatatlan halmazai lennének. Minél nagyobb az adathalmaz, annál nagyobb jelentősége van a jó minőségű, részletes metaadatoknak. A metaadatoknak saját logikájuk, rendszerük van, ami előzetes tervezést igényel.

A tervezést, az interoperabilitást és az adatcserét nagyban megkönnyítik, ha a metaadatok szabványosak. Az interoperabilitás (az a képesség, hogy az e-iratokat több szoftverrel és hardverrel lehet használni) azért fontos, mert ezzel csökkenthetjük a hard-

---

<sup>23</sup> Andrew McDonalds (éd.): *Les archives électroniques, Manuel à l'usage des archivistes* (ICA Studies 2005) 56 old. Elérhető: <http://www.ica.org/10801/studies-and-case-studies/ica-study-n16-electronic-records-a-workbook-for-archivists.html>

ver- és szoftverfüggést. Az adatcsere pedig azért, mert így lesz lehetséges adatok exportja és importja, adategyüttesek összeolvasztása, integrálása. A nemzeti metaadat-szabványok mellett hamar megjelentek a nemzetközileg elfogadottak is, amelyek közül a következőkben néhány fontosat említek. (A szabványokról bővebben a 8. fejezetben lesz szó.)

A Dublin Core (dublini mag) a nevét egy amerikai (ohio-i) kisvárosról kapta, mert 1995-ben ott ülésezett az a munkacsoport, amelyik azt kidolgozta, a mag-kifejezés pedig arra utal, hogy csak az alapvető, legfontosabb leíró metaadatokat tartalmazza.<sup>24</sup> Kezdetben egy 15 elemből álló Dublin Core Metadata Element Set-et dolgoztak ki, jelenleg ez az ún. egyszerű szint, a későbbiekben az elemek számát jelentősen bővítették és finomították, így született meg a kvalifikált szint. Az iratok leírására szolgáló egyszerű szint metaadat-elemei a következők: hozzájáruló (személy vagy szerv), kiterjedés (hely, idő, jogkör), létrehozó, idő, leírás, formátum, azonosító, nyelv, közreadó, kapcsolat, jogok, forrás, tárgy, cím, forrástípus.

A Dublin Core sosem vált népszerűvé a levéltárosi közösségben, mert nem veszi figyelembe a levéltári iratoknak az előző fejezetekben részletezett sajátosságait, ezért csak erős kompromisszumokkal lehet vele levéltári anyagot leírni. A fogalmai csak részben felelnek meg a levéltárakban alkalmazottaknak – a könyvtárak viszont jobban tudják használni –, de ami a legfontosabb, nem vesz tudomást az iratok hierarchikus struktúrájáról.

A nemzetközi levéltárosi közösség is gyorsan megalkotta a saját leíró metaadat-szabványait. (Az alább említendő szabványok részletesebb ismertetését lásd a 8.1, 8.4-8.5. fejezetekben!) Az első, a Nemzetközi Levéltári Tanács által 1994-ben publikált leírási szabvány az ISAD(g) (General International Standard Archival Description) volt, ami valószínűleg a szervezet eddigi legsikeresebb dokumentuma. Ezután két évvel, 1996-ban megjelent egy, az ISAD(g)-t kiegészítő másik leírási szabvány, az ISAAR(CPF) (International Standard Archival Authority Record for Corporate Bodies, Persons and Families). A harmadik nagyhatású nemzetközi szabvány, az EAD (Encoded Archival Description) kidolgozásához 1993-ban kezdtek a kaliforniai Berkeley Egyetemen, 1998-ban született meg az első verzió, majd 2002-ben a második. Az EAD már kimondottan informatikai alkalmazásra, XML leíró nyelvvel (lásd a 8.1. fejezetet!) készült – az ISAD(g) és az ISAAR ebből a szempontból semleges, nem operálnak informatikai fogalmakkal, így akár hagyományos segédletek készítésénél is lehet őket használni –, és bár az EAD-on kezdetben

---

<sup>24</sup> *Dublin Core Metadata Initiative*. Elérhető: <http://dublincore.org/>

erősen érezhető volt rajta az amerikai könyvtárosi hatás, a második kiadást viszont már igyekeztek összhangba hozni az ISAD(g)-vel. Viszonylag későn, 2011-ben megszületett az ISAAR(CPF) XML-es verziója is, az EAC (Encoded Archival Context).

Mivel az ISAD(g)-vel, az ISAAR-ral, az EAD-vel és az EAC-vel a későbbiekben részletesebben foglalkozom, most két további, a megőrzési metaadatokra vonatkozó nemzetközi szabványt ismertetek röviden.

A PREMIS (PREservation Metadata: Implementation Strategies) adatszótárt egy nemzetközi munkacsoport dolgozta ki kifejezetten a megőrzési metaadatok szabványosítása céljából, és elmondása szerint de facto nemzetközi szabvánnyá vált.<sup>25</sup> „Az adatszótár egy adatmodell köré szerveződik, ami öt entitásból áll, és amelyek a digitális megőrzési eljárással vannak kapcsolatban: intellektuális entitás, objektum, esemény, ágens.”<sup>26</sup> Az egyes entitások további egységekre, Entity Semantic Units-okra van bontva, pl. az Object Entity Object Identifier-re, Object Category-ra, Object Characteristics-re stb. A PREMIS felhasznál más szabványokat is, így kidolgozott egy csomó XML-sémát és igyekezett megfelelni egy eddig általam még nem említett szabványnak, az OAIS információs modellnek.<sup>27</sup>

A PREMIS alapján rögzített megőrzési metaadatokat az őrzéshez „be kell csomagolni” és össze kell kötni más leíró, adminisztratív és strukturális metaadatokkal. A METS (Metadata Encoding and Transmission Standard)<sup>28</sup>, ami egy olyan formát szabványosít, ami a „csomagolást” definiálja, ez lényegében nem más, mint egy OAIS Information Package XML implementációja (séma). Egy METS dokumentum hét szekciót foglal magába: fejléctet (a METS-dokumentumokat magukat írja le), leíró metaadatokat, adminisztratív metaadatokat, fájllistákat, strukturális leírásokat, strukturális kapcsolatokat és eszközöket (eszközök a tartalom és az applikációk összekötésére).

## 5 Az iratok integritása és lényeges jellemzői

Az integritás jelentése teljesség és koherencia. Egy hagyományos irat teljességét és koherenciáját viszonylag könnyen lehet ellenőrizni, az e-iratok viszont – amint ezt az előző fejezetekben láttuk – sokkal komplexebbek a hagyományosnál, az alkotó elemeik szétvál-

---

<sup>25</sup> PREMIS Data Dictionary for Preservation Metadata, version 2.2, July 2012. Elérhető: <http://www.loc.gov/standards/premis/v2/premis-2-2.pdf>

<sup>26</sup> PREMIS Data Dictionary for Preservation Metadata, 2.old.

<sup>27</sup> Lásd a 8.3. fejezetet!

<sup>28</sup> Elérhető: <http://www.loc.gov/standards/mets/>

nak, ezért az integritásukat sokkal nehezebb biztosítani, pedig az a hitelességük, a megbízhatóságuk és a használhatóságuk miatt alapvetően fontos. Ráadásul ez az integritás nemcsak az egyes iratok szintjén fontos, hanem az irategyütteseknél is.

A szakirodalom nem egységes abban, hogy mit tekint az e-irat integritásának. A teljességet mint követelményt többen<sup>29</sup> összekötik a fixitással. A fixitás: „a levéltári irat olyan minősége, ami egy tartós formát és stabil tartalmat garantál.”<sup>30</sup> Az integritás és a fixitás összekötése nem szerencsés, mert az utóbbit biztosítandó pl. meg kellene őrizni az eredeti szoftvert, hogy pont ugyanúgy nézzen ki az irat, mint eredetileg, illetve nem tudunk mit kezdeni a dinamikus (gyorsan, állandóan változó) iratokkal. Az integritás nem jelentheti azt, hogy a fájloknak és a metaadatoknak érintetlennek kell lenniük, ez a gyakorlatban – amint később meglátjuk – sokszor kivihetetlen.

Sokkal célravezetőbb, ha az integritást olyan teljességnek tekintjük, aminek az a feltétele, hogy az irat minden lényeges alkotó eleme meglegyen. Az előző fejezetekben leírtak alapján a hordozónak nem kell azonosnak lenni – mindegy, hogy egy e-iratot merevlemezről vagy CD-re mentett másolatáról töltjük-e be –, a formának csak részben (lásd alább!), a tartalomnak, az azonosítónak, a kontextusnak és a struktúrájának viszont meg kell őrződnie, egyébként főként ez utóbbi hordozza az irat koherenciáját. Az e-világban az iratok belső- és külső szerkezete erősen veszélyeztetett, különösen konvertálás esetén. Ekkor észrevétlenül eltűnhetnek olyan metaadatok (pl. az e-mailek címzettjei), melyek nélkül az irat értéke és hitelessége erősen csökken. Minél bonyolultabb egy irat vagy irategyüttes belső és külső szerkezete annál nehezebb feladat az integritásuk megőrzése és reprodukálása. Az adatbázisok, mint iratok, rendelkeznek a legbonyolultabb belső- és külső adatstruktúrával, még ha a felhasználó ezt nem is látja: egy, a képernyőn megjelenő lista adatait jellemzően több adattáblából rendezi össze a számítógép.

Az integritás kérdésével talán legmélyebben az InterPARES-projekt foglalkozott, ott a fogalom definíciójaként a következőt olvashatjuk: „Egy levéltári irat integritása a teljességre és az intakt kondíciójára utal: egy irat akkor integráns, ha a lényeges elemeiben teljes és változatlan.” Itt számunkra a lényeges elem a fontos: „Ez alatt az értendő, hogy a fizikai integritása, mint a bináris láncolatának pontos száma módosulhat, megváltoztatható azzal

---

<sup>29</sup> Így pl. Andrew McDonalds (éd.), *Les archives électroniques, Manuel à l'usage des archivistes* (ICA Studies 2005) 11. old. vagy Michael Day, *Instalment on "Metadata"* (DCC Digital Curation Manual 2005) 7. old.

<sup>30</sup> *Lignes directrices à l'intention des créateurs* (Projet InterPARES 2) 4. old. Elérhető: [http://www.interpares.org/ip2/display\\_file.cfm?doc=ip2\\_creator\\_guidelines\\_booklet\\_french.pdf](http://www.interpares.org/ip2/display_file.cfm?doc=ip2_creator_guidelines_booklet_french.pdf)

a feltétellel, hogy a tartalom szervezése és a szükséges annotációk, valamint a dokumentációs formák változatlanok maradjanak.”<sup>31</sup>

Az Ausztrál Nemzeti Levéltár szerint az irat lényegét (record essence) kell megőrizni: ilyen például egy szövegszerkesztett iratnál a tartalom, a betűszedés, a grafikák stb., de nem tartoznak ide például a toolbar-ok vagy a képernyő eszköztára (gombok, színek stb.). Az ausztrál levéltárosok elismerik, hogy „az irat lényege nem egy tudomány és eléggé függ a szubjektív megítéléstől...” Az irat lényegét irattípusonként (szövegszerkesztett irat, e-mail stb.) állapítják meg, hogy aztán az ellenőrzés – hogy a lényeg megvan-e – már formális mechanizmus legyen.<sup>32</sup>

Az ausztrál koncepció levéltártanon alapszik, mások inkább az irattani vagy a technikai szempontokat hangsúlyozzák. Az, hogy mi a lényeges, az ugyanis nemcsak irattípusonként, hanem fájl típusonként is eltérő lehet, hiszen egy szövegfájlnak mások a technikai specifikációi, mint egy videófájlnak. Ezt feltérképezendő indultak olyan projektek (pl. az InSPECT vagy a PLANETS<sup>33</sup>), amelyek különböző típusú digitális objektumok azon lényeges jellegzetességeit (significant properties) definiálták, amelyek megléte szerintük biztosítja az objektum integritását. (Erre alapozva lehet aztán olyan szoftvereket kifejleszteni, amelyek automatikusan ellenőrzik ezeknek a jellegzetességeknek a meglétét pl. konverzió után.) Az InSPECT pl. a strukturált szövegeknél 37 jellegzetességet állapított meg, amelyben vannak a tartalomra (pl. cím, létrehozó, kulcsszavak) vagy formára (pl. sortörés, aláhúzott, félkövér) utaló információk és egyéb adatok (pl. gomb, listaelemek).<sup>34</sup>

Láthatjuk, hogy a nemzetközi szakirodalom rendkívül sokat foglalkozik az egyes iratok integritásával, azaz a belső struktúrával, de meglepően keveset a külső struktúrával, azaz irategyüttesek integritásával. Pedig ez is fontos lenne, mert amint az előzőekben láttuk az egyes levéltári iratok nem állnak meg önmagukban, a jelentésük és értékük csak a többi irattal együtt érvényesül, tehát a külső integritást is meg kell őrizni.

A külső integritást másképpen kell garantálni, mint a belsőt. Az iratképzőknél keletkezett iratok nagy részét – úgy az analógokat, mint az e-iratokat – ki kell selejtezni, ez pedig

---

<sup>31</sup> *Conditions requises pour évaluer et maintenir l'authenticité des documents d'archives électroniques* (InterPARES) 3. old. Elérhető: [http://www.interpares.org/display\\_file.cfm?doc=ip1\\_authenticity\\_requirements%28french%29.pdf](http://www.interpares.org/display_file.cfm?doc=ip1_authenticity_requirements%28french%29.pdf). Meg kell jegyezni, hogy a dokumentációs formát illetően a nézetünk eltérő.

<sup>32</sup> Helen Heslop & Simon Davis & Andrew Wilson: *An Approach to the Preservation of Digital Records* (National Archives of Australia, December 2002) 13-15. old. Elérhető: [http://www.naa.gov.au/Images/an-approach-green-paper\\_tcm16-47161.pdf](http://www.naa.gov.au/Images/an-approach-green-paper_tcm16-47161.pdf)

<sup>33</sup> <http://www.planets-project.eu/>

<sup>34</sup> Stephen Grace & Gareth Knight & Lynne Montague: *InSPECT Final Report* (2009) 20. old. Elérhető: <http://www.significantproperties.org.uk/inspect-finalreport.pdf>

az integritás ellen dolgozik. Viszont ha a selejtezést megfelelő módszerrel (minél kevesebb kapcsolódást elvágva), az irattári tervvel összhangban végezzük, és megfelelően dokumentáljuk az eljárást (tehát leírjuk, hogy mit selejteztünk ki), akkor nemcsak a hitelességet, de a megfelelő külső integritást is biztosítjuk. Ugyanakkor tudomásul kell venni, hogy vannak olyan iratok, amelyek külső integritása – bármennyire is szeretnénk – nem őrizhető meg. Ilyenek pl. a weblapok, amelyek jellemzően tele vannak linkekkel (más weblapokra ugrás lehetőségével): egyrészt a weblapok nagyon gyorsan változnak (dinamikusak), másrészt a linkekkel további lapokra mutatnak, tehát ha követnénk azokat, akkor az archiválás mértani hatványként növekedne.

Külön gondot jelenthetnek az adatbázisok, nemcsak a bonyolult struktúrájuk miatt, hanem mert 1) sokuk dinamikus, azaz a tartalmuk gyorsan változik, 2) a hosszú ideig működő adatbázisok szerkezetben és funkcióban is változhatnak, 3) az adatbázisok jellemzője az átmeneti – néha tranzakciónak nevezett – rekordok léte, amit a használó hoz létre pl. kereséskor, szűréskor. Ezekre a problémákra nincs univerzális megoldás, sokkal inkább kompromisszumokban kell gondolkodni. Az egyik lehetőség az ún. pillanatfelvételek létrehozása, amikor elmentjük és archiváljuk egy adatbázis állományát egy adott időben. Ennek az a nagy hátránya, hogy egy dinamikus adatállomány statikus lenyomatát hozzuk létre, ráadásul ezt időről-időre meg kell ismételni, ami hosszú távon és sok adatbázissal számolva óriási tárhelyet követel. Jobban kíméli a tárhelyet az ún. inkrementális mentés, amikor – egy teljes mentés után – csak a megváltozott rekordokat mentjük el, így lehet később rekonstruálni a pontos változásokat. Ez viszont nem segít akkor, ha lényegesen változott az adatbázis szerkezete vagy funkciója. Az átmeneti rekordokat nem érdemes megőrizni, viszont az adatbázis azon funkcióit, eszközeit, amelyek ilyen rekordok létrehozását lehetővé teszik, feltétlenül.<sup>35</sup>

Amint az előzőekben láttuk, az integritáshoz hozzá tartozik a kapcsolódó iratok (irategyüttesek) összetartozásának megőrzése is. Bár sok szervnél külön gyűjtik pl. az e-maileket, ezek többnyire egy-egy ügghöz kapcsolódnak és más iratokkal (levelekkel, memorandumokkal stb.) együtt alkotnak egy-egy ügyiratot. A szerveknél egy ügyben egyaránt keletkeznek analóg (papír) és e-iratok – ezeket nevezik hibrid iratoknak –, az összetartozásukat az irattári rendszernek kell biztosítani. A levéltárakba pedig hasonlóképpen sok vegyes irat kerül, sőt a meglévő papír- vagy mikrofilmezett iratok egy részét is digitali-

---

<sup>35</sup> A vázolt kérdésekről lásd pl. Luciana Duranti & Kenneth Thibodeau: The Concept of Record in Interactive, Experiential and Dynamic Environments: the View of InterPARES. In: *Archival Science* 6, 1: 13-68. old. (2006) vagy Malcom Todd: *File Formats for Preservation* (Technology Watch Report, Digital Preservation Coalition 2009) 27-29. old. Elérhető: <http://www.dpconline.org/advice/technology-watch-reports>

zálják. A külső integritást a levéltárban a különböző szintű nyilvántartásoknak és segédleteknek kell biztosítaniuk, ezeket pedig főként leírás révén készítik el a levéltárosok.<sup>36</sup>

## 6 Az iratok hitelessége

A hitelesség már az analóg másolatoknál is problémát okozott, pedig ezek némelyike, például a mikrofilm, bizonyos tekintetben szinte hamisíthatatlan volt (pl. egy felvételen belül nem lehet manipulálni egy irat szövegét). A hitelességnek különböző fokai vannak a hagyományos iratoknál. Egy adásvételi szerződésnek szigorú formai és tartalmi szabályai vannak, Magyarországon például nem elég csak az aláírás, tanúk is kellene, sőt, minden oldalt minden aláírónak a kézjegyével kell ellátnia. Egy egyszerű, minden irattárban előforduló emlékeztetőnél már nincsenek pontos szabályok, de ajánlatos a készítőjének aláírnia, különben korántsem biztos, hogy hitelesnek bizonyul. Mindezek ellenére a levéltárban őrzött iratok igen nagy százaléka nincs aláírással ellátva, ez azonban nem jelenti azt, hogy ezek egyáltalán nem tekinthetők hitelesnek. Vannak közismert technikák (például pecsételés, jelzetelés), amelyek növelik a hitelességet, ezeken kívül azonban léteznek más, bonyolultabb hitelerősítő tényezők is: pl. a szerv iratkezelését pontosan rögzítő szabályzat, irattári terv, átadókönyvek, tételjegyzékek stb. mind ezt szolgálják.

Az e-iratok hitelességét leginkább az teszi kétségessé, hogy könnyen manipulálhatók, és a beavatkozásnak közvetlenül nem marad nyoma: egy szövegfájlban például könnyedén átírhatunk egy mondatot, egy képfájlban átszínezhetünk, dúsíthatunk egy részt, egy hangfelvételtől eltüntethetjük a háttérzajt. Ráadásul, amint arra az előzőkben többször utaltam, az irat alkotó elemei elválhatnak és rendkívül sérülékenyek, a hordozó gyakran változik, a forma módosulhat (pl. konvertáláskor).

A hitelesség nem azonos a jogi bizonyító erővel, annál szélesebb fogalom, amennyiben a társadalmi ítéletre utal: a használóknak (kutatóknak) is hitelesnek kell elfogadniuk az iratot. Közkeletű definíció szerint egy irat akkor hiteles, ha valóban az, aminek állítja magát. A hitelességet gyakran összekötik más fogalmakkal, mint a megbízhatóság (egy irat képessége, hogy hiteles bizonyítékként szolgáljon), az identitás (azok az attribútumok, amelyek karakterizálják és megkülönböztetik más iratoktól), az integritás és a fixitás.

A hitelesség kérdésével legrészletesebben az InterPARES-projekt foglalkozott, szerinte két szinten kell biztosítani a hitelességet:<sup>37</sup>

---

<sup>36</sup> Erről bővebben lásd a 8.4. és a 9. fejezeteket is.

1) Az iratok szintjén az identitása és integritása által.

2) Az iratképző és -őrző szintjén a következők által

- hozzáférési korlátozások az iratképzésnél, a módosításnál, az annotációnál és a selejtezésnél;
- védelmi eljárások az iratok elvesztése és változtatása ellen és technológiai támogatás;
- dokumentációs forma minden egyes eljáráshoz;
- a megfelelő iratok kiselejtezése vagy átadása a levéltárnak.

A Nemzetközi Levéltári Tanács kézikönyve együtt tárgyalja a megbízhatóság, a hitelesség, az azonosság, az integritás, valamint a használhatóság és az eltarthatóság követelményeit. Ezeket a feltételeket szerinte akkor lehet biztosítani, ha az iratok keletkezésekor, használatakor, levéltári átvételekor, feldolgozásakor és őrzésekor

- az iratok képzése, az információhoz való hozzáférés, az adatok változtatása és törlése megfelelően kiépített jogosultsági rendszer alapján történik, és az eljárások megfelelően dokumentáltak (pl. naplózás);
- az iratoknak egyedi azonosítójuk van;
- minden irathoz megfelelő metaadat tartozik, ami kapcsolatot jelent az irat és a munkafolyamat, valamint az irat és más kapcsolódó iratok között;
- a metaadat az irat keletkezésétől kezdődően dokumentálja az irat sorsát (például rögzít minden formátumváltást), és az irathoz hasonlóan védett (azaz nem változtatható);
- az iratokat rendszeresen ellenőrzik, hogy a szükséges konvertálás kellő időben megtörténjék;
- minden irat automatikusan vagy csekély beavatkozással exportálható nyílt (open) őrzési formátumokba anélkül, hogy lényeges tartalom-, struktúra- vagy kontextusvesztés történne;
- az iratokhoz megfelelő dokumentáció tartozik, amiből kiolvashatók a kontextusok és a struktúra;

---

<sup>37</sup> *Conditions requises pour évaluer et maintenir l'authenticité des documents d'archives électroniques* (InterPARES) 6-9. old. Elérhető: [http://www.interpares.org/display\\_file.cfm?doc=ip1\\_authenticity\\_requirements%28french%29.pdf](http://www.interpares.org/display_file.cfm?doc=ip1_authenticity_requirements%28french%29.pdf).

- az iratok megfelelően védettek megsemmisülés és károsodás ellen.<sup>38</sup>

A fenti felsorolás egy sor olyan technikai, irattári eljárásra utal, amelyek szükségesek a hitelesség megőrzéséhez, ilyen pl. a check sum, az események naplózása, ami a felhasználótól függetlenül történik, aki azt nem tudja megváltoztatni, a metaadatok rögzítése és az elektronikus aláírás. Ez utóbbiról érdemes egy kicsit bővebben beszélni.

Mint ismeretes, a világon az ún. kétkulcsos digitális aláírás terjedt el, aminek az a lényege, hogy minden jogosult személynek két kulcsa van: egy író kulcs, ezzel „írja alá” a tulajdonos az iratot, illetve egy olvasó kulcs, ezzel lehet ellenőrizni, hogy valóban az illető személy írta-e alá a kérdéses iratot. Mindkét kulcs tulajdonképpen egy-egy szoftver, az írókulcshoz hardver, token vagy csippel ellátott kártya is tartozik. Az írókulcs egy szigorúan titkos, jogosultanként változó kód szerint elemzi az iratot, majd az elemzés eredményeként egy kódot generál, amit az irathoz rendel („aláírja”). Mivel minden egyes irat eltérő – kivételek a másolt fájlok – ezért az „aláírás” iratonként eltérő lesz. Az olvasó kulcs, ami nyilvános, két szempontból vizsgálja az „aláírt” iratot: 1) tartalmazza-e a jogosult személy kulcsát a generált kód, 2) az iratot az „aláírás” után megváltoztatták-e (ha igen, akkor a kód nem „passzol”). Ha az első kérdésre a válasz igen, a másodikra pedig nem, akkor az irat „aláírása” hiteles.

A digitális aláírás sem nem megfajthatatlan, sem nem örökkévaló. Becslések szerint az informatika fejlődése olyan gyors, hogy az aláíró algoritmusok 5-10 éven belül feltörhetőek lesznek, tehát maximum ilyen időközönként szükségessé válik új kulcsok kibocsátása, illetve a már aláírt dokumentumok újrakódolása. Az aláírások ellenőrzése, az állandó frissítés a levéltárak számára megoldhatatlan, nem is lennének jogosultak arra, de ha jobban belegondolunk, akkor ez nem is az ő feladatuk, ők csak azért vállalnak felelősséget, hogy az átvétel óta hitelesen őrzik az iratokat, erre pedig a garanciát maguk az iratok, az átvételi, a kezelési és az őrzési körülmények adják, mint ahogy arról az előbbieken szó volt. Érvényes ez akkor is, ha az elektronikusan aláírt iratokat a többivel együtt konvertálják, tehát megváltoztatják. Mindazonáltal a levéltárak politikája nem teljesen egységes e tekintetben. Pl. a Public Record Office Victoria (Ausztrália) konvertálás után az iratokat „összecsomagolja” a metaadatokkal, majd ezt a csomagot saját maga digitálisan aláírja, és minden jövőbeni változást újabb csomagolás és aláírás követ.<sup>39</sup> Az InterPARES az e-aláírások

---

<sup>38</sup> Andrew McDonalds (éd.): *Les archives électroniques, Manuel à l'usage des archivistes* (ICA Studies 2005) 30. old. A felsorolás az ISO 15489-1-es szabványára épül.

<sup>39</sup> Michael Day: *Instalment on "Metadata"* (DCC Digital Curation Manual 2005) 19.old. Elérhető: <http://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/metadata/metadata.pdf>

kiküszöbölését javasolja és – mint az amerikai és az ausztrál Nemzeti Levéltár – az aláírásokra vonatkozó információk és metaadatok megőrzését tanácsolja, így dokumentálva az digitális aláírás tényét.<sup>40</sup>

Amikor biztosítjuk, hogy egy irat vagy irategyüttes hiteles, megbízható, integráns és fix legyen, akkor elérjük, hogy az – az adminisztratív/praktikus, bizonyító és történelmi – értékek érvényesüljenek, az életciklusuktól függően különböző mértékben, ahogy az 1.2. fejezetben láttuk. Ha ez nem sikerül, akkor az értékek kisebb vagy nagyobb mértékben sérülnek, így pl. nem segítik a munkafolyamatot a szerveknél, nem szolgálnak bizonyítékként egy jogvitában, vagy nem megbízhatóak a geológus vagy a történész számára.

## 7 A fájlok (megőrzési) formátumai

A fájlformátumok kérdése az egyik legfontosabb a hosszú távú megőrzés biztosítását illetően. A fájlformátum egy adott fájl előzetes, specifikus adatstruktúrája (adatrendezése és -szervezése). A szoftvereket egy vagy több, meghatározott formátum olvasására, illetve kezelésére fejlesztik ki, és ha az olvasandó fájl nem tartozik ezek közé, akkor vagy csak részlegesen lesz olvasható, vagy egyáltalán nem. Egy fájl nem biztos, hogy csak egy formátumot takar: az XML- és SGML-formátumok (8.1. fejezet) például egyszerű szövegfájl bázisúak, lehetnek egyszerű szövegfájlok, de gyakori eset, hogy egy XML-es fájlban más formátumú fájlok vannak „becsomagolva”.

A formátumok az informatikai ipar termékei, és meglehetősen sok van belőlük, valamint változékonyak (bár nem annyira, mint a szoftverek vagy hardverek). Más típusú formátumok kellenek az állóképekhez, mozgóképekhez, hang- és szöveges dokumentumokhoz stb., új formátumok jelennek meg és régiak tűnnek el, a meglévőket tovább fejlesztik, egy formátumnak több verziója is lehet forgalomban. Mindez természetesen a hosszú távú megőrzés ellen dolgozik. A formátumok között minőségi különbségek is vannak, pl. képet vagy hangot az egyik jobban reprodukál, mint a másik, nem mindegy, hogy integrálja-e vagy sem a metaadatokat, több-e a helyigénye vagy kevesebb, komplex-e vagy egyszerű.

Az 1990-es években még általános volt az a nézet, hogy a hosszú távú megőrzésre csak bizonyos formátumok alkalmasak – hogy melyek, arról persze voltak viták –, és az őrző intézménynek ragaszkodnia kell ezekhez a „kanonizált” formátumokhoz. Az azóta

---

<sup>40</sup> *Lignes directrices à l'intention des créateurs*, Projet InterPARES 2, 7. old. Elérhető: [http://www.interpares.org/ip2/display\\_file.cfm?doc=ip2\\_creator\\_guidelines\\_booklet\\_french.pdf](http://www.interpares.org/ip2/display_file.cfm?doc=ip2_creator_guidelines_booklet_french.pdf), valamint <http://www.archivesdefrance.culture.gouv.fr/static/1051>, 3. old.

eltelt két évtizedben két ok miatt változott ez a szemlélet. Egyrészt korábban nagyszerűnek tartott formátumokról kiderült, hogy több tekintetben hátrányosak, másrészt nyilvánvalóvá vált, hogy az őrző intézmény funkciójával, stratégiájával összhangban kell kiválasztani a formátumokat, amelyek intézményenként eltérők lehetnek. Pl. a Holland Nemzeti Könyvtár az 1990-es években a PDF-formátumot preferálta, mivel azzal a bekerülő dokumentumokat jól lehetett reprezentálni. Az idők folyamán azonban egyre változatosabb anyagok kerültek be, pl. weblapok, médiafájlok, adatbázisok, ezért a könyvtár felülvizsgálta a formátumpolitikáját, és ma már sokkal több formátumot fogad be.<sup>41</sup> A British Library 80 terabyte-nyi, korábban TIFF-formátumban digitalizált újságot konvertált át JPEG2000 formátumba, mert ezzel jelentősen csökkenteni tudta az őrzési költséget.<sup>42</sup>

Az biztos, hogy a levéltárak hosszú távon nem tudnak túlságosan sok formátumot befogadni, mert ezzel éppen úgy járnának, mint a sok és elavuló szoftverrel. Az átveendő fájlformátumok számának csökkentése – ha úgy tetszik: intézményi szabványosítása – nagy dilemma: ha mindent vagy nagyon sokféle formátumot elfogadnak, akkor előbb vagy utóbb megjelenítési és használati gonddal néznek szembe. Ha erősen redukálnak, akkor a konvertálások miatt, akár a jelenben is, megjelenítési (minőségi) gondok merülhetnek fel.

A formátumok számának korlátozását három módon lehet elérni:

- Az iratképzők elektronikus iratkezelési rendszere már eleve a követelt fájlformátumokat produkálja – erre jogszabályban kell kötelezni őket.
- Az iratképzők többféle fájlformátummal dolgoznak, de a levéltárnak már csak az előírt formátumokra konvertálva adják át az iratokat – ez is megfelelő jogszabályi háttértételez.
- Az iratképzők többféle fájlformátumban adják át az iratokat a levéltárnak, az egységsítés (csökkentés) konvertálás útján a levéltárban történik.

A levéltár szempontjából a legjobb megoldás az első, mert így az iratok lényegében változtatás nélkül kerülnek át, a második és harmadik hátránya a konvertálás, ami óhatatlan információ-vesztéssel jár. (A konvertálások elkerülhetetlenek, de a számukat a lehető legkevesebbre kell szorítani.) A legelőnytelenebb a harmadik megoldás, mert a konvertálás technikailag bonyolult és rendkívül költséges feladata a levéltár költségvetését és sze-

---

<sup>41</sup> Judith Rog & Caroline van Wijk: *Evaluating File Formats for Long-term Preservation*, 1. old.. Elérhető: [http://www.kb.nl/sites/default/files/docs/KB\\_file\\_format\\_evaluation\\_method\\_27022008.pdf](http://www.kb.nl/sites/default/files/docs/KB_file_format_evaluation_method_27022008.pdf)

<sup>42</sup> *Preservation and Long Term Access via NETWORKED SERVICES*, Planets Project, 14. old. Elérhető: <http://www.planets-project.eu/>

mélyzetét terheli. (A levéltárnak a későbbi konvertálások során már egy viszonylag egységes formátumú adattömeget kell átalakítania, ami egyszerűbb feladat.)

Így vagy úgy, a megfelelő formátumpolitika kidolgozása alapvetően fontos minden intézménynek, amely hosszú távon kíván megőrizni e-iratokat. Vannak intézmények, amelyek csak általánosságban fogalmazzák meg a követelményeiket, mások konkrétan előírják, hogy milyen formátumokat tudnak fogadni. A nemzetközi szakirodalomban sok követelmény- vagy szempontrendszer olvashatunk, az alábbi kritériumok majdnem mindenütt szerepelnek:<sup>43</sup>

- Mennyire elterjedt a formátum. A nagyobb elterjedtség előny, mert több szoftver tudja használni, és általában jobb eszközök (pl. konvertálók) állnak rendelkezésre.
- Mennyire független a formátum egy speciális hardver/szoftvertől. A nagyobb függetlenség természetesen előny.
- A formátum specifikációi mennyire megismerhetők (publikusak). A könnyen hozzáférhető (publikált) és jól dokumentált (leírt) formátum előnyös, mert jobban megítélhető és jobb alkalmazások (programok) írhatók rá.
- Mennyire azonosítható és ellenőrizhető a formátum identitása és attribútumai, mennyire akadályozza ezt pl. tömörítés, wrapper, titkosítás, beleértve a digitális aláírást is. Minél transzparensőbb egy fájl, minél inkább automatizálható az azonosítás, annál jobb.
- Mennyire van ellátva a formátum metaadatokkal és azok mennyire nyerhetők ki (metadata support). Természetesen a gazdag metaadat-ellátottság és a könnyű ki-nyerhetőség előnyös.

A fenti kritériumok mellett számosan hangsúlyozzák annak a fontosságát, hogy a formátumnak

- újrahasználhatónak és interoperábilisnak kell lennie, minél több hardver és szoftver képes kezelni a formátumot, annál jobb;
- egyszerűnek kell lennie – feltéve, hogy a tartalmat jól meg tudja jeleníteni –, és ellenállóknak a hibákkal szemben, azaz belső hibakorrekciós lehetőségekkel kell rendelkeznie (robustness/complexity/viability);

---

<sup>43</sup> Az alábbi felsorolást lásd Malcom Todd: *File Formats for Preservation* (Technology Watch Report, Digital Preservation Coalition 2009) c. művéből vettem (13-15. old.). Elérhető: <http://www.dpconline.org/advice/technology-watch-reports>

- ha fejlesztési ciklus része, akkor visszamenőleges kompatibilitást kell biztosítani (stability);
- lehetőleg nem szabad jogilag védettnek lennie (intellectual property), mert az akadályozza a szabad használatot és költséges is lehet.

A kritériumok az egyes formátumoknál keveredhetnek. A PDF pl. elterjedt, de jogvédett, ugyanakkor publikált, lehet komplex és egyszerű, strukturált és nem strukturált. A Microsoft DOC-fájlok rendkívül elterjedtek, jogvédettek de nem (illetve csak részlegesen) publikáltak.

A sok millió iratot tároló őrző intézményeknek tekintettel kell lenniük a tárhelyigényre is. Egyes iratok, főként a képek, de méginkább a mozgóképek fájlok nagyon terjedelmesek. A tárolási igény csökkentésére és az adatátvitel könnyítésére számos tömörítési eljárás született, amelyek alapvetően két csoportra oszthatók: 1) adatvesztés nélküli tömörítési eljárásokra, ilyen például az LZW (Lemple-Zif-Welch) vagy az RLE (Run Length Encoding) eljárás, amelyek pl. a TIFF-formátumú képek helyigényét képesek csökkenteni; 2) adatvesztéses tömörítési eljárásra, közülük a legismertebbek a különböző fokozatú JPEG-tömörítések (Joint Photographics Experts Group). Az adatvesztés nélküli tömörítések jellemzően csak kevésbé csökkentik a fájl méretét, viszont, – mint ahogy az elnevezésük is mutatja – a fájl minősége nem károsodik. Az adatvesztéses tömörítéseknél a rendkívül jó méretcsökkenés árát pl. a képek finomabb részleteinek elvesztésével kell megfizetni. Ha viszont ismételten tömörítünk, akkor már komoly veszteség állhat elő.

Az 1990-es években az őrző intézményekben még általános volt a tömörítésektől való idegenkedés, a digitális iratok robbanásszerű növekedése viszont óhatatlanul kompromisszumra kényszerített több intézményt. A londoni The National Archives pl. a kiadott útmutatójában (2008) elfogadhatónak tartja a veszteséges tömörítést, a mester- (tehát a napi használatban nem használt) példányoknál is, ha a minőségromlás a képeknél vizuálisan nem észlelhető. Ugyancsak tolerálja a veszteséges tömörítést olyan e-iratok esetében, ahol rendelkezésre áll a papír eredeti is.<sup>44</sup>

Nincs egyöntetű vélemény a komplex vs. egyszerű formátumokat illetően. Az egyszerű formátumban kevesebb a hibalehetőség, ami hosszú távon – pl. konvertálások során – csökkenti a fájl sérülését vagy hozzáférhetetlenségét, viszont a bonyolult e-iratok informá-

---

<sup>44</sup> Lásd *Image Compression* (Digital Preservation Guidance, note 5) 10. old. Elérhető: <http://www.nationalarchives.gov.uk/documents/information-management/image-compression.pdf>, illetve *Digitization at The National Archives* (2013), 9. old. Elérhető: <http://www.nationalarchives.gov.uk/documents/information-management/digitisation-at-the-national-archives.pdf>

cióit csak komplex formátum tudja jól reprezentálni, ennek hiányában romlik a minőség. Tehát ebben a tekintetben is mérlegelni kell, és szükség esetén kompromisszumra kell jutni.

Mindazonáltal sok levéltár és könyvtár kiadja azon formátumok listáját, amelyeket hosszú távon alkalmazni kíván. Vannak olyan intézmények, amelyek szigorúan korlátozzák az elfogadott formátumok számát, ilyen pl. a svájci Archives fédérales, illetve vannak olyanok, amelyek sokkal megengedőbbek, ilyen pl. a londoni vagy a washingtoni National Archives. A teljesség igénye nélkül, csupán tájékoztatás céljából érdemes felsorolni, hogy napjainkban az egyes irattípusoknál melyek a népszerű, és több tekintélyes intézmény által hosszú távú őrzésre elfogadott vagy ajánlott formátumok:

- Strukturálatlan szövegek: PDF/A, TXT
- Strukturált szövegek: XML, ODT, PDF/A
- Táblázatok: XML, ODS, CSV
- Adatbázisok: XML, CSV
- Állóképek: TIFF, JPEG2000, PNG
- Audio: WAV, WMA
- Video: MXF, MOV, AVI
- E-mail: XML, MBOX, EML

Nem egyszerű dolog a formátumok azonosítása és ellenőrzése. A fájlkiterjesztés (pl. .TIFF, .PDF), ami elvileg jelzi, hogy milyen formátumról van szó, önmagában nem jelenti azt, hogy a fájl az, aminek hisszük. Sokféle formátum-variáció létezik a világban és sok nem felel meg a publikált struktúrának, specifikációknak. Ez azért lehet nagy baj, mert ha a fájl olvasásakor – amikor a szoftverünk próbálja használni a fájlt – vagy konvertálásakor derül ki az inkompatibilitás, az elérhetetlenné teheti az e-iratot vagy akár egy egész adatbázist. Az őrző intézmény csak úgy alkalmazhat migrálási stratégiát (lásd a 9.1. fejezetet!), ha a rendszerébe kerülő fájlokat azonosítja és ellenőrzi még konvertálás előtt. Ennek a munkának fontos eszközei a formátumregiszterek. Több van ilyen a világban, az alábbiakban a brit Nemzeti Levéltár által eredetileg belső használatra kifejlesztett, ma már az interneten bárki által elérhető PRONOM-regisztert ismertetem röviden.<sup>45</sup>

---

<sup>45</sup> Lásd: <http://www.nationalarchives.gov.uk/aboutapps/pronom/#documentation>

A PRONOM jelenleg több mint 800 fájlformátum pontos leírását és technikai specifikációit tartalmazza, valamint információkat azokról a szoftverekről, amelyek támogatják az egyes formátumokat. A PRONOM része a DROID, egy szoftver, amellyel megvizsgálhatjuk a fájlaink technikai jellemzőit, és azonosíthatjuk őket. Amint az előzőekben említettem, a formátumok az informatikai ipar termékei, megvan az életciklusuk, létrehozzák őket, továbbfejlesztik, variációk készülnek belőlük, leállnak a fejlesztéssel, elavulnak. Ezért fontos, hogy az olyan regiszterek, mint a PRONOM, karban legyenek tartva, állandóan kiegészüljenek, amit a brit Nemzeti Levéltár jelenleg vállal.

## 8 A szabványok

A 2.1. fejezetben láthattuk, hogy a levéltáros szakmai közösséget milyen nagy kihívások érték az elmúlt 130 évben, és hogy ezekre a közösség milyen szakmai, azaz elméleti és gyakorlati válaszokat adott. Az e-iratok gyors térhódításáig, kb. a 1990-es évek elejéig, a levéltártan szorosan kapcsolódott a történettudományhoz, és a szakmai fejlődés nagy mérföldköveit egy-egy nagyhatású elméleti munka jelentette – mint pl. a fentebb említett holland kézikönyv vagy az amerikai Theodore R. Schellenberg munkái –, hasonlóképpen a történettudományhoz. Az e-világban viszont a levéltártudomány – logikus módon – sokat átvett az információtudománytól, és ez a közeledés abban is megnyilvánul, hogy elmondhatjuk: a szabványok váltak az új archivisztika mérföldköveivé.

A szabványoknak azért van kiemelt jelentőségük az informatikában, mert a számítógép absztraháló képessége rendkívül korlátozott – igaz, fejlődik –, és a szabályostól (a beprogramozottól) való kis eltérés is azonnal hibát eredményez. Ezért nem ismerik fel a szoftverek az „idegen” formátumot, ezért nem tudnak együttműködni egymással „idegen” programok. A megoldás a szabványosítás, ez ellen viszont sok minden dolgozik.

Az informatikai világ érdekes dichotómiája a proliferáció (burjánzás) és az egységesülés párhuzamos jelenléte, illetve egyensúlya hosszú távon. Erre a világra már a kezdetektől fogva, de különösen az internet általánossá válása óta jellemző volt a szabadság, aminek egyik hajtóereje az a tény, hogy az e-információt könnyű létrehozni, változtatni, sokszorosítani, kombinálni és terjeszteni. A túlzott szabadság – hardver, szoftver, hálózat, formátum stb. tekintetében – azonban anarchiát okoz, akadályozza a megértést és a terjesztést, ezért egy bizonyos ponton automatikusan megindult az egységesülés, ami főként jogalkotásban és szabványok elfogadásában jelentkezik.

A szabványok lehetnek kanonizáltak – azaz egy illetékes nemzetközi vagy nemzeti hatóság által elfogadottak és kihirdetettek –, ilyen pl. a JPEG2000, PDF/A, TIFF, ODF formátumok. Lehetnek de factók – azaz hatóságok által nem elfogadott, de a közösségben igen széles körben használtak –, ilyen pl. a Microsoft Office formátumai. A szabványokat nevezhetik szabványnak, modellnek vagy követelményeknek, ahhoz azonban, hogy egy szabványnak tényleges hatása legyen, mindenképpen de factónak kell lennie, akár kanonizált, akár nem.

A szabványok is az e-világ produktumai, ezért rájuk is jellemző az előbb említett dichotómia, azaz egyrészt a burjánzás – bármennyire is ellentmond ez az eredeti céljuknak –, mert egyre több van belőlük, egymástól gyakran eltérőek, párhuzamosak, illetve másrészt az egységesülés, ami esetünkben egybeolvadást, integrálódást, egymásra épülést jelent. Néhány példa: az 1990-es évek elejéig számtalan nemzeti szabvány létezett a levéltári leírás egységesítésére (azaz leírási metaadatokra), majd a Nemzetközi Levéltári Tanácsgondozásában megjelent az ISAD(g), ami óriási hatással bírt a nemzeti szabványokra is, és sok közülük a későbbiekben be is építette az elemeit. Az ISAD(g) megjelenése után 2 évvel megjelent az ISAAR, ami az ISAD(g) egyik információs területét, az iratképzőre vonatkozó adatokat részletezte, ezzel lehetővé téve, hogy a proveniencia elvet nem követő levéltárak is alkalmazhassák az ISAD(g)-t.<sup>46</sup> Ezzel párhuzamosan Amerikában kidolgozták az EAD-t, ami – az ISAD(g)-vel és ISAAR-ral ellentétben – csak e-környezetben használható, hiszen XML leíró nyelvet használ.<sup>47</sup> Míg az EAD első verziója erősen kötődött a Library of Congress eredetileg könyvtári alkalmazásra kifejlesztett MARC-formátumához, addig a 2002-ben publikált második verzió már bevallottan is eltávolodott tőle, és közelített az ISAD(g)-hez, tulajdonképpen integrálta azt. Hasonlóan jártak el az ISAAR-ral, amikor az EAD pandanjaként 2011-ben létrehozták az EAC-t. Hasonló átfedések, illesztések és párhuzamosságok figyelhetők meg OAIS, a METS és a PREMIS szabványok között.<sup>48</sup>

A szabványoknak is megvan a rövidebb vagy hosszabb élettartamuk, rájuk is jellemző a változás, ami talán ellentmondásnak tűnik, de érthető is, hiszen az e-világ alapvető tulajdonsága nemcsak a változás, hanem a fejlődés is. A hosszú élettartamú szabványoknak gyakran újabb és újabb verziói jelennek meg, és bár ezek mindig az előző verziókra épül-

---

<sup>46</sup> Az ISAD(g)-t és az ISAAR-t illetően lásd a 8.4. fejezetet.

<sup>47</sup> Az XML ismertetését lásd a 8.1., az EAD-t a 8.5. fejezetben.

<sup>48</sup> A METS-re és a PREMIS-re – lévén metaadat-szabványok – a 4. fejezetben már hivatkoztam, az OAIS ismertetésére a 8.3. fejezetben került sor.

nek, az új kiadások mégis a szabvány egyik fontos funkciója, az időtállóság ellen dolgoznak.

A szabványok között a siker mérőfoka egyrészt az élettartam hosszúsága, másrészt az, hogy hányan alkalmazzák, illetve milyen más szabványok alkalmazkodnak hozzá. Az egyik legsikeresebb az XML leíró nyelv, ami megállíthatatlanul terjed az 1990-es évek óta. A 4. fejezetben említett Dublin Core szabványt levéltári körökben először az ISAD(g)-ISAAR, majd később az EAD-EAC teljesen háttérbe szorította. Feltétlen sikertörténetnek tekinthetjük a MOREQ-et,<sup>49</sup> ezt az európai iratkezelési modellt, valamint az OAIS-t, ami az iratok-adatok hosszú távú megőrzésére szóló szabvány, ezek számtalan más szabványba épültek be, használták fel. Az OAIS hivatkozási ponttá vált a megőrzési közösségben, amennyiben OAIS-kompatibilissá teszik a szabványaikat vagy az ajánásaikat. A következő alfejezetekben majd ezeket a sikeres szabványokat ismertetem röviden.

Az kétségtelen dolog, hogy az őrző intézményeknek a gyarapítási tevékenységükhöz, az adatcseréhez, de különösen a hosszú távú megőrzéshez nélkülözhetetlenek a szabványok. Azokat gondosan kell megválasztani, és pedig két fő szempontot tartva szem előtt: 1) a használhatóságot, azaz, hogy a szabvány a célnak megfelel-e és jól alkalmazható-e; 2) biztonságot, azaz hogy hosszú távon alkalmazható-e és széles körben elfogadott-e, illetve van-e szervezet a karbantartásra, fejlesztésre (a gyakran változó szabványok hátrányosak). Tehát tudomásul kell venni, hogy a szabványok sem örök életűek, az életük során is változnak, és hosszú távon elkerülhetetlen az új szabványra való áttérés, ami óhatatlanul konvertálást (migrálást) jelent, ez pedig adatvesztést eredményezhet.

## 8.1 XML

Az XML<sup>50</sup> (Extensible Markup Language, azaz kiterjesztett kijelölő nyelv) egy nyelvcsalád tagjaként született. Elsőként az SGML-t (Standard Generalized Markup Language, azaz szabványos, általános kijelölő nyelv) fejlesztették ki, 1986-ban lett szabvány (ISO 8879), később ebből alakították ki a HTML-t (Hypertext Markup Language, azaz hipertext kijelölő nyelv), majd az XML-t 1998-ban. Ezt a nyelvcsaládot szoftverfüggetlennek is nevezik, mert egyszerű szövegbázisú, azaz a szintaktikája szabad szemmel is olvasható.

Az XML-t elsősorban strukturált szövegek feldolgozására fejlesztették ki, a struktúrát, illetve annak elemeit pedig tag-ekkel lehet reprezentálni. A tag-ek mindig csúcsos záróje-

---

<sup>49</sup> Ismertetését lásd a 8.2. fejezetben!

<sup>50</sup> Az XML gondozója, a World Wide Web Consortium (W3C). Lásd: <http://www.w3.org/standards/xml/>

lek (< >) között vannak, így különülnek el a szövegtől, amit strukturálnak. Mindig párosan vannak, van egy nyitó és van egy záró tag, megelőzik, illetve lezárják azt a szövegrészt amire vonatkoznak. A záró tag formailag majdnem megegyezik a nyitóval, csak „/” jellel kezdődik. Nézzünk egy példát a jelen mű egy korábbi szövegrészből:

```
<title>4. A metaadatok</title>
```

```
<para>A metaadatok a közkeletű definíció szerint adatok az adatokról (iratokról). A metaadatokat az analóg világ is ismerte, csak nem így hívta őket...</para>
```

A példában a <title> </title> tag-ek határolják a címet, a <para> </para> tag-ek pedig a bekezdést.

A tag-eken kívül még vannak elemek, attribútumok (értékkel), kommentárok stb., melyek segítségével nagyon bonyolult iratokat lehet reprezentálni.

Az XML egy önleíró nyelv, ami azt jelenti, hogy a tag-ekben szereplő adatelemeket és attribútumokat, valamint a közöttük lévő kapcsolatokat (azaz a struktúrájukat) az alkalmazónak kell definiálnia (leírnia) egy ún. a DTD-ben (document type definition) vagy sémában, ez az XML legfontosabb része. Az XML-lel a formát is meg lehet jeleníteni, az irathoz külön készíthetünk stíluslapot (stylesheet). Előny, hogy egy irathoz különféle célokra különböző stíluslapokat készíthetünk, ezáltal különféleképpen jeleníthetjük meg azt anélkül, hogy magát a szöveget tartalmazó fájlt módosítanánk.

Természetesen az XML szerint készült e-iratok nem manuálisan készülnek, és bár az előbb hangsúlyoztuk, hogy a szintaktika szabad szemmel is olvasható, az adatbevitelt, ellenőrzést, olvasást, kezelést természetesen automatikusan, megfelelő szoftverrel kell végezni.

Az XML szinte forradalmat indított el az informatikai világban, főleg a sokoldalú alkalmazhatósága, egyszerűsége és időtállósága miatt. A különböző szakmák egymás után hozták létre a saját hatáskörükbe tartozó DTD-eket, illetve sémákat, így a levéltárosi közösség is, nemzeti és nemzetközi szinten egyaránt, az utóbbira a legjobb példa az EAD és az EAC (8.5. fejezet).

## 8.2 MOREQ

Egy általános (európai) elektronikus iratkezelési követelményrendszer kidolgozására a DLM-Fórum<sup>51</sup> kapott megbízást az Európai Bizottságtól még az 1990-es években. Ennek eredményeként született meg 2001-ben a Model Requirements for the Management of Electronic Records, ami rövidítve a MOREQ nevet kapta. Az azóta eltelt közel másfél évtizedben a MOREQnek már több új verziója is megjelent, és olyan sikeres elektronikus iratkezelési szabvánnyá vált, hogy már Európán kívül is alkalmazzák. A MOREQ fontos forrás lehet a kormányzatoknak, amelyek egységesíteni kívánják a közszféra iratkezelését, de a szoftverfejlesztők számára is, akik ez alapján tervezhetik meg a fejlesztendő iratkezelő programot és rendszert.

Fölmerülhet a kérdés, hogy az iratok hosszú távú megőrzését tárgyaló műben milyen helye van egy olyan szabványnak, ami az iratoknak csak az irattári életciklusával foglalkozik, tehát addig, amíg az iratok a levéltárba nem kerülnek. Azt azonban tudni kell, hogy a levéltári életciklus szempontjából alapvetően fontos a szervnél eltöltött irattári ciklus, ugyanis sok minden ott dől el. Ha az iratképző az iratokat rosszul sorolja be, nem fordít gondot az összetartozó iratok egyben tartására, sok az elkeveredett, elveszett irat, rosszul végzi a selejtezést stb., akkor a levéltár már olyan rossz minőségű anyagot fog kapni, amit nem tud érdemben javítani. Az e-iratok esetében ez ugyanígy van – lényegében ugyanazokat a feladatokat kell elvégezni, mint a papíriratoknál –, sőt talán még fontosabb, mert a beavatkozás egy rossz e-rendszerben rendkívül nehéz, ha nem lehetetlen, pl. nem lehet manuálisan, egyenként metaadatolni több tízezer e-iratot. Ideális esetben a levéltáros véleményét már az iratok születése előtt, az e-iratkezelő rendszer kialakításánál kikérik.

A legújabb verzió, a MOREQ2010,<sup>52</sup> mint az elődei, szintén az ISO 15489-es iratkezelési szabványból indul ki, viszont annál lényegesen bővebb. A funkcionális követelményeket kilenc „szolgáltatás”-ra építi, pl. felhasználói és csoport alkalmazás, besorolási alkalmazás, selejtezési alkalmazás, keresői alkalmazás.

Ezen kívül egy sor nem funkcionális követelményt részleteznek, pl. a biztonsággal, a magánélettel, a megbízhatósággal stb. kapcsolatban.

A MOREQ2010 adatformátumként az XML-t ajánlja, és kifejlesztettek egy XML sémát, ami előfeltétele lehet az exportnak és importnak.

---

<sup>51</sup> A DLM Fórumot (DLM = Document Lifecycle Management) az Európai Bizottság hozta létre 1994-ben. A honlapja: <http://www.dlmforum.eu/>

<sup>52</sup> Elérhető: [http://www.moreq.info/index.php?option=com\\_content&view=article&id=17&Itemid=22&lang=e](http://www.moreq.info/index.php?option=com_content&view=article&id=17&Itemid=22&lang=e)

### 8.3 OAIS

Az OAIS (Open Archival Information System) egy referencia modell nyitott információs archiváló rendszerek számára<sup>53</sup>, amit a Consultative Committee for Space Data Systems (CCSDS) nevű űrkutatási szervezet dolgozott ki (ISO szabvánnyá [14721] vált 2002-ben). Az e-információk hosszú távú levéltári megőrzését és az adatok kommunikációját szolgálja.

Az OAIS modellje az információkat csomagokban kezeli, egy csomag megfelelő tartalomtól és hozzá tartozó metaadatokból áll, azaz tulajdonképpen egy logikai egység. A különböző eljárásokhoz (pl. beszállítás, megőrzés, kommunikáció) különböző csomagok tartoznak. Három információs csomag van::

- Beszállítandó információs csomag (SIP): az a csomag, amit a beszállító küld a levéltárnak.
- Archivált információs csomag (AIP): az a csomag, amit a levéltár megőrzés céljából készít a SIP(ek)ből, ez a legfontosabb (conservation).
- Terjesztési információs csomag (DIP): olyan csomag, amit a felhasználó kap, amikor a kért információkat lekérdezi (communication).

A három csomag információtartalma lehet nagyon hasonló. Az OAIS azonban nemcsak adatmodellt kínál, hanem funkcionális modellt is, pontosan leírja, hogy egy őrző levéltárnak milyen feladatai vannak, milyen entitásokat kínál, és azokhoz milyen funkciók tartoznak, illetve hogy azokhoz milyen csomagokat kell kialakítani. Pl. belépési entitást, raktározási entitást, adatkezelési entitást, hozzáférési entitást stb.

### 8.4 ISAD(g) és SAAR(CPF)

A Nemzetközi Levéltári Tanács (ICA) 1994-ben publikálta a Levéltári leírás általános nemzetközi szabványát (angol rövidítése: ISAD(g)).<sup>54</sup> Kétségtelen, hogy ez az egyik legfontosabb dokumentum, amit az ICA valaha is kiadott. Az ISAD(g) nemcsak témaként vagy idézetként tűnt fel azóta sok-sok publikációban, hanem az egész világon jelentősen befolyásolta a levéltári gyakorlatot. Sok levéltár az ISAD(g)-t figyelembe véve vizsgálta felül a leírási módszereit és adatbázis-rendszerét.

---

<sup>53</sup> Elérhető: [http://pin.association-aristote.fr/lib/exe/fetch.php/public/documents/norme\\_oais\\_version\\_francaise.pdf](http://pin.association-aristote.fr/lib/exe/fetch.php/public/documents/norme_oais_version_francaise.pdf)

<sup>54</sup> Az ISAD(g) második kiadása 2000-ben jelent meg. Elérhető: [http://www.ica.org/10225/normes/ISAD\(g\)-norme-gnralt-et-internationale-de-description-archivistique-deuxieme-dition.html](http://www.ica.org/10225/normes/ISAD(g)-norme-gnralt-et-internationale-de-description-archivistique-deuxieme-dition.html)

A levéltári anyag leírása, ami jellemzően iratcsoportok (fond, állag, sorozat stb.) leírása, különösen fontos, mert ezzel nemcsak az anyagban lévő információt tesszük elérhetővé, hanem rögzítjük (hitelesítjük) is azt és ezzel együtt az anyag rendszerét is.<sup>55</sup>

Az ISAD(g) 26 adatelemet 7 információs területre csoportosít:

- Azonosító adatok (azonosító, cím, keletkezési időpont, a leírás szintje [pl. fond], terjedelem [pl. ifm]).
- Kontextus (az iratképző szervezet vagy személy neve, szervezeti vagy életrajzi ismeretetés, az iratok története, a levéltárba kerülés jogcíme).
- Tartalom és struktúra (kör [pl. időkör] és tartalom, az iratértékelésre és a selejtezésre vonatkozó információk, gyarapodásokra [pl. újabb beszállítások] vonatkozó információk, az iratanyag rendszere).
- A kutathatóság és a használat feltételei (a használattal és a reprodukálással kapcsolatos feltételek, az iratok nyelve, fizikai jellemzők és technikai követelmények, a rendelkezésre álló segédletek).
- Kapcsolódó anyagok (az eredetiek őrzési helye; másolatok megléte és helye; kapcsolódó irategyüttesek; az iratokra vonatkozó publikációk).
- Megjegyzés.
- A leírásra vonatkozó információk.

Az ISAD(g) koncepciója az iratszinteken alapul, önmagát többszintű leírásnak (multilevel description) hívja, az általa leírt iratanyagnál pontosan be kell mutatni a külső kapcsolatokat. Fontos ajánlása, hogy a levéltári iratokat felülről lefelé, azaz a nagyobb irategyüttesektől a kisebbek felé haladva kell leírni (tehát először a fondot, majd az állagait, aztán az állagok sorozatait stb.). Minden szinten csak az ott releváns információt kell megadni, ezzel elkerülhetővé válik az informatikai alkalmazásoknál oly' gyakran elkövetett hiba: az adat- vagy információismétlés.

Az ISAD(g) szigorúan a provenienciaelven alapszik, figyelembe veszi a levéltári információ/anyag törvényszerűségeit, és kiegyensúlyozott leírást ad az iratok négy alkotó komponenséről, az azonosítóról, tartalomról, a struktúráról és a kontextusról.

---

<sup>55</sup> Feltételezve, hogy a leírás követi az anyag rendjét. Előfordul, hogy a leírás (levéltári segédlet) nem követi az eredeti rendet, pl. amikor szétszórt fondokat egyesítünk egy e-leírásban.

Az ISAD(g) megjelenése után két évvel (1996-ban) az ICA kiadott egy másik leírási szabványt, A személyek és szervek, kollektivitások, családok leírásának nemzetközi szabványát (angol rövidítése: ISAAR[CPF]).<sup>56</sup> Ez a szabvány nem kínál komplett leírást, részben az ISAD(g)-re épül, és a két szabvány együttes alkalmazását javasolják. Az ISAAR tulajdonképpen az ISAD(g) kontextus információs területe egy részének – az iratképző szervezetre vagy személyre vonatkozó információk – részletes kifejtése.<sup>57</sup> (31 adatelemet 5 információs területre csoportosít.)

## 8.5 EAD és EAC

Az 1980-as években az amerikai levéltárak többnyire a könyvtárakban kifejlesztett MARC (Machine Readable Cataloging) formátumban írták le elektronikusan az irataikat.<sup>58</sup> Sok levéltáros azonban elégedetlen volt ezzel, riasztották őket a könyvtárak rendkívül erősen szabványosított fogalmai, a rengeteg információelem, amiket nem éreztek alkalmasnak a sokszínű levéltári anyagának leírásához.<sup>59</sup> A legsúlyosabb kifogásuk az volt, hogy egy MARC-rekord egy gyűjtemény szintű információnak felel meg a levéltárakban.<sup>60</sup> Az amerikai levéltárosi közösségben így hamarosan megfogalmazódott egy új szabvány megalkotásának szükségessége, amely

- nyílt kódú formátumra épül (bárki által megismerhető, ezáltal felhasználói alkalmazások – célszoftverek – írhatók az adatok kezelésére),

---

<sup>56</sup> Az ISAAR második kiadása 2004-ben jelent meg. Elérhető: <http://www.ica.org/10230/normes/isaar-cpf-norme-internationale-sur-les-notices-dautorit-utilises-pour-les-archives-relatives-aux-collectivites-aux-personnes-ou-aux-familles-2me-dition.html>

<sup>57</sup> Fölmerül a kérdés, hogy miért nem az ISAD(g) megfelelő információs területét bővítették, miért kellett létrehozni egy új szabványt. Erre csak egy logikus magyarázat van: lehetővé kellett tenni olyan levéltári anyag (fond, állag) leírását, amelynek több iratképzője van. Ez viszont ellenkezik a provenienciával.

<sup>58</sup> Lásd: [www.loc.gov/marc/](http://www.loc.gov/marc/). Pontosabban szólva a levéltárak számára adaptált változatot használták. Az amerikai levéltárak rendkívül erősen kötődnek a könyvtárakhoz – az EAD is a washingtoni Kongresszusi Könyvtár honlapján elérhető el –, és ennek történeti okai vannak. Az amerikai levéltárügy viszonylag későn szerveződött meg (a National Archives-t például csak 1934-ben alapították, kutatók, történészek kitaró lobbizása után), azelőtt az európai értelemben vett levéltári iratokat vagy maguk a szervek vagy könyvtárak őrizték. Az ok valószínűleg az amerikai társadalom-felfogásra és gazdasági filozófiára vezethető vissza. Eszerint az ember eredendően szabad lény, az emberek, a közösségek önmagukért felelősek, és a lehető legnagyobb lehetőséget kell biztosítani a szerveződésükhöz, illetve működésükhöz. Ez gazdasági szempontból többé-kevésbé a „laissez faire, laissez passer”-elv érvényesülését, az állam tekintetében pedig a lehető legkisebb állami beavatkozást, bürokráciát eredményezte. Az 1929-ben kezdődött nagy gazdasági válság azonban alapjaiban rengette meg ezt a felfogást, és innentől (pontosabban szólva a New Deal-től) datálhatjuk az erős központi (szövetségi) intézményrendszer kiépülésének kezdetét, valamint az állami beavatkozás általános erősödését. Ez viszont automatikusan nagyobb bürokráciát, illetve nagyságrendekkel nagyobb tömegű iratot eredményezett, ami óhatatlanul maga után vonta a levéltári intézményrendszer gyors kiépülésének szükségességét. Mindezek ellenére az USA-ban a könyvtári befolyás még ma is nagyon erős a levéltárakra. Mind a köz-, mind a magánkönyvtárak rengeteg levéltári anyaggal rendelkeznek, bevett gyakorlat, hogy jelentős közéletli személyek vagy utódaik nem levéltárnak, hanem könyvtárnak adják át az irathagyatékukat.

<sup>59</sup> Helen R. Tibbo & Lokman I. Meho: Finding Finding Aids on the World Wide Web. In: *American Archivist*, vol. 64. Spring 2001, 63. old.)

<sup>60</sup> EAD Application Guidelines for Version 1.0 (<http://www.loc.gov/ead/ag/agcontxt.html>)

- szoftverfüggetlen,
- megfelelően leírja a hagyományos levéltári, könyvtári és múzeumi segédleteket,
- megjeleníti az egymással (hierarchikusan vagy nem hierarchikusan) összefüggő levéltári információkat, és lehetővé teszi a szintek közötti mozgást,
- támogatja az elemspecifikus indexálást,
- valamint lehetővé teszi a rendszerek közötti interaktív kapcsolatokat (adatlekérdezés, adatátöltés, adatcsere).

Nyilvánvaló volt, hogy az újonnan kidolgozandó szabványhoz – ami az EAD (Encoded Archival Description, azaz Kódolt levéltári leírás) nevet kapta –, az adatformátumot (keretet) a piacon lévők közül kell kiválasztani, hiszen a levéltárak is részei az információs társadalomnak, így esett a választás az SGML leíró nyelvre, majd az EAD második verziójánál (2002) az XML-re.<sup>61</sup> A második verzió abban is újdonságot hozott, hogy erősen közeledett az ISAD(g)-hez, az adatelemeit úgy alakította át, hogy azok megfeleljenek annak.

Az EAD tulajdonképpen egy SGML/XML-szintaxist alkalmazó levéltári DTD (lásd a 8.1. fejezetet!). Az elmúlt évtizedben de facto szabvánnyá vált, azaz a levéltárosi közösség elfogadta és alkalmazza.

Az EAD jóval több adatelemmel (146-al) operál, mint az ISAD(g) és az ISAAR (26, illetve 31), ráadásul mindehhez járul még 101 attribútum, ami jól mutatja, hogy jóval részletesebb, mint a másik kettő. Nézzünk meg néhányat az elemek közül:

- Descriptive Identification <did>: a levéltári anyag azonosítója.
- Title of the Unit <unittitle>: a levéltári anyag címe.
- Date of the Unit <unitdate>: a levéltári anyag időköre.
- Extent <extent>: a levéltári anyag fizikai terjedelme.
- Abstract <abstract>: a levéltári anyag tartalma.

Most lássunk néhány attribútumot:

- Audience: a leírt adathoz való hozzáférők köre.
- Inclusive: az adat (például dátum) bezárólagosan értendő.
- Level: a leírandó levéltári anyag szintje (pl. fond, állag, sorozat stb.).

---

<sup>61</sup> Az EAD első és második kiadása elérhető: <http://www.loc.gov/ead/>

- Numeration: a számozás karakterei (pl. arab, római stb.).

Az ISAD(g)-vel és az ISAAR-ral összehasonlítva az EAD lényeges eltérése, hogy más oldalról közelíti meg a levéltári leírás problematikáját. A leírás termékét elektronikus iratnak tekinti, és választ kíván adni az e-iratokkal kapcsolatos kérdésekre. Az EAD nagy erénye, hogy a leírás (viszonylag) szoftverfüggetlen, időtálló, könnyen konvertálható, szabványos adatelemekkel bír, modularizálható, és többféle felhasználást tesz lehetővé. Ezzel a szabvánnyal lehetőségessé válik, hogy minden létező iratanyagot leírjunk, a meglévő segédleteinket pedig átírjuk, azaz összes leírásunkat egy nagy keretbe foglaljuk, ezáltal jelentősen könnyebbé válik a keresés, a böngészés, illetve bármilyen adatkarbantartás.

Viszonylag későn, 2010-ben született meg az ISAAR XML-es pandanja, az Encoded Archival Context (Kódolt levéltári kontextus), Corporate Bodies, Persons, and Families (EAC-CPF), ami egy XML-es séma, tehát informatikai nyelvre fordította az eredeti szabványt<sup>62</sup>. Ugyanúgy, mint az ISAAR-t, ezt sem áll meg önmagában, hanem célszerű az EAD-vel használni.

## 9 Az elektronikus iratok tartós megőrzésének stratégiái

Az előző fejezetekben részletesen kifejtettem, hogy milyen nehézségek merülnek fel az e-iratok hosszú távú megőrzésénél: hardver-, szoftver- és formátumavulás, szabványok hiánya vagy nem alkalmazása, az e-irat bonyolult szerkezete és a szerkezet sérülékenysége, az alkotó elemek szétválása, a könnyű manipulálhatóság, szerzői jogi problémák, az iratok átmeneti vagy dinamikus jellege. A későbbiekben majd említek még további nehézségeket, mint pl. vírusok, titkosítás vagy a védettség (pl. jelszóval). Ezek ismeretében, de akár az e-iratokkal kapcsolatos pár évtizedes tapasztalatainkból is manapság már egyértelműen levonhatjuk azt a következtetést, hogy az e-iratokat nem tudjuk változatlan formában megőrizni évtizedeken, évszázadokon keresztül – illetve ha így cselekszünk, akkor az iratok egy idő után olvashatatlanok és használhatatlanok lesznek –, a legtöbb, amit elérhetünk, hogy az iratokat olyan állapotban tartjuk, vagy olyan állapotba hozzuk, hogy a belátható jövőben képesek leszünk azokat reprodukálni.<sup>63</sup> Minden, hosszú távú megőrzésre berendezkedett intézménynek ki kell alakítania a megőrzési stratégiáját, ami tulajdonkép-

---

<sup>62</sup> AZ EAC teljes (angol) verziója elérhető: <http://eac.staatsbibliothek-berlin.de/>

<sup>63</sup> Kenneth Thibodeau szavai, idézi Terry Eastwood: The Appraisal of Electronic Records: What is New? In: *Comma* 2002/1-2. 77. old.

pen azt jelenti, hogy meghatározza azokat az alapvető módszereket, amelyek alkalmazásával ez a reprodukálási képesség megtartható.

A megőrzési stratégia sikerének feltétele, hogy a levéltár megfelelő minőségű, rendezett iratokat kapjon a szervektől (lásd a 8.2. fejezetet!). A rendezettség azt jelenti, hogy az iratok egy előre elhatározott szempontrendszer, az irattári terv szerint be vannak sorolva a megfelelő kategóriákba. Rendezetlen iratokat – legyenek azok analógok vagy elektronikusak – nem lehet értékelni és selejtezni, azok egy áttekinthetetlen, hozzáférhetetlen halmazmazzá válnak – még akkor is, ha egyenként olvashatók –, és ez olyan, mintha az iratok elpusztultak volna.

A megőrzés fontos feltétele, hogy a levéltárba került iratok ott kellően fel legyenek dolgozva: levéltári rendszerbe kell őket foglalni, azaz állagokba, fondokba és fondcsoportokba sorolni, ezt a hagyományos iratoknál a fond- és állagjegyzék jeleníti meg, e-iratoknál ugyanezt általában adatbázis formájában látjuk. (A levéltári rendszerbe sorolás és az iratok leírása általában egyidejűleg történik.)

A megőrzési stratégiát illeszteni kell az őrző intézmény egyéb stratégiáihoz, pl. a gyűjtési politikájához, (milyen anyagokat gyűjt, azoknak mik a sajátosságai), vagy a terjesztési és kommunikációs politikájához (kik fogják használni az iratokat, milyen céllal, milyen mértékben teszik hozzáférhetővé pl. a neten). A hosszú távú megőrzés végső célja, hogy a használók (a kutatók) rendelkezésére álljanak az iratok, akik ki tudják aknázni azok értékeit. A rendelkezésre állás elérhetőséget jelent, az e-iratok tekintetében pedig feltétlenül azt, hogy a használók on-line tudják olvasni az iratokat, az pedig megfelelő hardvert, szoftvert, adatbázist, formátumokat stb. követel meg. A levéltárak és könyvtárak kommunikációs forradalma ma már valóság, és óriási mértékben kiterjesztette a kutatótermet, az internet révén az egész glóbuszra. Sőt, már megindult a világ könyvtárainak és levéltárának „egyesítése” is, amikor integrált portálokon, egységes felületen érhetőek el az iratok digitális képei. (Lásd pl. az Europeana- vagy az Apenet-projektet.)<sup>64</sup>

Érdemes kockázatelemzést végezni, hogy milyen veszélyek fogják fenyegetni az iratokat, és természetesen fel kell mérni a várható költségeket és tárhelyigényeket, a szakemberszükségletet is. Sőt a stratégiák sem örök életűek, 5-10 évente érdemes elemezni a követett stratégia eredményességét – ennyi idő alatt nagyot változhat az informatikai világ –, és megtenni a szükséges korrekciókat, ami eredményezhet stratégiaváltást is, már ha a körülmények azt megengedik.

---

<sup>64</sup> Lásd <http://www.europeana.eu/portal/>, illetve <http://www.archivesportaleurope.net/>

A levéltárak az elmúlt pár évtizedben sokféle megőrzési stratégiával próbálkoztak, voltak köztük sikeresek, azaz időtállóak, ezeket a következő fejezetekben fogjuk ismertetni, és voltak sikertelenek is, amelyek közül kettőt azért érdemes röviden felidézni.

Az egyik az ún. technológiai megőrzés, aminek az a lényege, hogy nemcsak az eredeti fájlokat őrzi meg a levéltár, hanem a kezelő szoftvert és hardvert is. Ez a stratégia napjainkban teljesen visszaszorult, ugyanis még egy kis gyűjtőkörű levéltár esetében is megoldhatatlan gondot jelent az állandóan növekvő gép- és szoftverpark üzemeltetése és karbantartása. Pár évtized után biztosan ellehetetlenülne a levéltár, mert az elavult hardverelemek gyártását megszüntetik, így a gépek (és velük együtt az általuk kezelt iratok) addig élnek, amíg az alkatrészeik működőképesek. Megoldhatatlan gondot jelentene a több száz felhalmozott szoftver kezelése, karbantartása is. Mindezek ellenére bizonyos esetekben van létjogosultsága a technológiai megőrzésnek. Érdekes olyan elavult hardverelemeket megőrizni 1-2 évtizedig, pl. számítógépeket floppylemez-meghajtóval, amelyek használata szükséges lehet a későbbiekben. Így az esetleg előkerülő régi lemezek egy ideig olvashatók, illetve átmenthetőek lesznek mai, szabványos hordozókra.

A másik, az ún. poszt-kustodiális stratégia az 1990-es években, az internet robbanás-szerű terjedésének kezdetén volt népszerű téma a levéltáros konferenciákon. Eszerint a jövő levéltára nem fog elektronikus iratokat őrizni és feldolgozni, mert azok a szervek őrzésében maradnak – a világháló idején közömbös, hogy melyik központi számítógépen vannak fizikailag az iratok –, és ők maguk fogják karbantartani, szükség esetén konvertálni az adatokat (iratokat). A levéltárak feladata csak az lesz, hogy az internet segítségével intellektuális kontrollt érvényesítsenek az iratok fölött, ide értve a leírást is (elsősorban a kontextus kibontására helyezve a hangsúlyt).<sup>65</sup> A poszt-kustodiális stratégia nem terjedt el, mert az iratképzők – és mivel zömmel közszereplőkről lévén szó –, az állam nem volt érdekelt abban, hogy pénzt és fáradságot nem kímélve évtizedekig, évszázadokig őrizzen és kezeljen régi iratokat. (És nem mellesleg ez nem is lett volna a feladata.)

Végül meg kell említeni egy olyan megoldást, ami részben emlékeztet a poszt-kustodiális stratégiára. A felhőben való tárolásról van szó, ez viszonylag új megoldás, nincs kellő tapasztalat vele, de eléggé foglalkoztatja az intézményeket. Igazából legfeljebb

---

<sup>65</sup> Az elmélet rövid összefoglalóját lásd: Terry Cook: Archives in the Post-Custodial World; Interaction of Archival Theory and Practice since the Publication of the Dutch Manual in 1898. In: *Archivum* XLIII, 205-207. old. Ilyen stratégiát követett az ausztrál National Archives az 1990-es években. [http://www.naa.gov.au/Images/an-approach-green-paper\\_tcm16-47161.pdf](http://www.naa.gov.au/Images/an-approach-green-paper_tcm16-47161.pdf), 6. old.

részstratégiának nevezhetjük, mert csak az e-iratok tárolására és technikai működtetésére vonatkozik, a megőrzés viszont sokkal szélesebb spektrumot átfogó politika.

A felhőben való tárolás azt jelenti, hogy nem az illetékes levéltárban, hanem attól távol egy másik intézmény vagy vállalat szerverén és kezelésében történik a tárolás és a szükséges szoftverek működtetése, az iratok elérése pedig a világhálón keresztül történik. Ilyen, tipikus felhőszolgáltatást végez a Google a Gmail-lel. A felhő lehet több szempontból előnyös – pl. költségkímélő, a levéltárnak nincs gondja infrastruktúrára (hardver, szoftver), nem kell informatikai szakembereket alkalmazni stb. –, azonban van egy óriási kockázata: az iratok kikerülnek a levéltár kontrollja alól. Minden szakmai munkát – stratégia kidolgozása, visszakereső rendszer kialakítása, regisztrálás, ellenőrzés, validálás, selejtezés, leírás stb. – a levéltárnak kell végeznie, csak a műszaki, informatikai feladatok egy részét bízta a tárolást végző cégre. Bár a fejlett országokban nagy a felhőszolgáltatások biztonsága, mégis alaposan mérlegelnie kell minden intézménynek, amely ilyen stratégiát fontolgat, hogy az előnyök arányban vannak-e a kockázattal. Ha mégis a felhő mellett dönt, akkor számos biztosítékot kell beépíteni a szolgáltatásba, ilyen pl. a megfelelő exportfunkció (hogy az adatok könnyen és biztonságosan kinyerhetők legyenek a rendszerből) vagy a többszörös másolat készítése.

## 9.1 Migráció

Az e-iratok archiválását végző levéltárak többsége a migrálási stratégiát alkalmazza. A migrálás azt jelenti, hogy a levéltár a fájlokat a régi, elavult formátumokról új szabványoknak megfelelő formátumokra konvertálja. Az új formátum kiválasztásánál a legfontosabb szempont a minőség és a tartósság, valamint mindazok a szempontok, amelyeket a 7. fejezetben felvázoltam. Mivel a formátumok is változnak, ezért a migrálást újra meg újra el kell végezni.

A migrálás során elkerülhetetlen az adatvesztés, de nem mindegy, hogy milyen mennyiségű és minőségű adatot veszítünk. A tartalmat, a metadatokat, illetve a lényeges funkciókat reprezentáló adatokat nem vagy csak minimális mértékben érheti veszteség, a formai megjelenítéssel kapcsolatos hiányosságnál már lehetünk nagyvonalúbbak. Az iratok integritásának ismertetésénél (5. fejezet) már említettem az ausztrál levéltárosok definícióját az irat lényegéről (record's essence), amit szerintük mindenképpen meg kell tartani. A konzekutív migrálás – azaz amikor már migrált anyagot újra migrálunk – azonban halmozott adatvesztést eredményez, ezért érdemes azt vagy elkerülni, vagy a lehető legritkábban, de az adatbiztonságot még nem veszélyeztető időintervallumonként csinálni. Az

adatvesztés ellenőrzésére nem elég a szűrőpróbaszerű ellenőrzés, a konvertálást végző szoftverbe olyan hibafigyelőt kell építeni, ami azonnal jelzi, ha egy iratot nem sikerült maradéktalanul átmenteni.

A konszekutív migrálás elkerülésének bevált módszere, ha a konvertálásokat mindig vagy a megőrzött eredeti fájlokról, vagy hosszú életű megőrzési fájlokról végezzük. Ilyenkor három rendszer állhat a rendelkezésünkre: 1) az eredeti fájlok, 2) megőrzési fájlok, 3) használati fájlok. Az eredeti fájlok megőrzése a biztonságot szolgálja, mert ha nem sikerül a konvertálás, akkor még mindig újra lehet végezni a műveletet. Ilyen politikát folytat az Egyesült Királyság Nemzeti Levéltára, de ugyanezt ajánlja a CEDAR-projekt is.<sup>66</sup> Az amerikai National Archives kifejlesztett egy Digital Preservation Software Platformot, amelyen keresztül érkeznek az új e-iratok (beszállítások). A platform legfontosabb eleme a XENA szoftver (XML Electronic Normalising for Archives), amellyel megőrzési fájlknál alkalmazott formátumokra konvertálják a bejövő fájlokat. Ez tehát egy nagyfokú szabványosítás, amit ők normalizálásnak neveznek. Hasonló stratégiát folytat az ausztrál Public Record Office, Victoria, amely kifejlesztette a Victorian Electronic Records Strategy-t (VERS), ami „hosszú távú” formátumokra – ezt neveztem fentebb megőrzési fájlformátumnak – konvertálja a bejövő iratokat. (A canberrai Nemzeti Levéltárhoz hasonlóan az ő szabványosításuk is XML-bázisú.)<sup>67</sup>

A migrálást érdemes a lehető leghamarabb, mindjárt a beszállítás után elvégezni, mert ekkor még általában kellő információ és szakismeret áll rendelkezésre mind a piacon, mind az iratképzőnél, ha esetleg a konvertálás nem sikerülne és korrigálni kell. Nem szabad vírussal fertőzött, titkosított, vagy (jelszóval) védett iratokat átvenni és természetesen konvertálni sem, mert az a későbbiekben is hozzáférhetetlen lesz.

A már migrált állományt is érdemes időnként ellenőrizni, és ahogy említettem, hosszú távon – ahogy változik az informatikai környezet – elkerülhetetlen az újbóli konvertálás. Ezt tervezetten, előre elhatározott periódusonként vagy nagyobb változások előtt – például tízévenként vagy az operációs rendszer változtatása előtt – kell végezni, nem pedig kampányszerűen, amikor az a veszély fenyeget, hogy az elavult formátumokat már nem is tud-

---

<sup>66</sup> Lásd <http://www.nationalarchives.gov.uk/aboutapps/pronom/#documentation>, valamint David Holdsworth: *Instalment on "Preservation Strategies on Digital Libraries"* (DCC Digital Curation Manual 2007) 12-13.old. Elérhető: <http://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/preservation-strategies/preservation-strategies.pdf>

<sup>67</sup> Lásd: [http://www.naa.gov.au/Images/Digital-Preservation-Software-Platform\\_tcm16-60788.pdf](http://www.naa.gov.au/Images/Digital-Preservation-Software-Platform_tcm16-60788.pdf), illetve <http://prov.vic.gov.au/government/vers>

juk maradéktalanul konvertálni. A legjobb megoldás persze az, ha olyan hosszú életű formátumokat alkalmazunk, amelyek nem avulnak el 10 évenként.

A migrálással létrehozott egységes formátumok jó lehetőséget teremtenek arra, hogy a levéltár kialakítsa az egész digitális állományára érvényes egységes kereső rendszerét, ami nagyban megkönnyíti a levéltárosok és a kutatók munkáját.

## 9.2 Emuláció

Az emulálás azt jelenti, hogy az eredeti informatikai környezetet (a hardvert, az operációs rendszert és a szükséges segédsoftvereket) külön erre a célra készült szoftverrel (emulátorral) imitáljuk, így illesztve azt jelenlegi platformhoz. Így lehetővé válik, hogy az adatokat az eredeti alkalmazói programmal egy teljesen más (új) hardver- és szoftverkörnyezetben is úgy futtassuk, mintha az eredeti környezetben tennénk ezt.

Az emulálás komoly informatikai felkészültséget igényel, mert az eredeti elektronikus iratok, valamint az alkalmazói szoftver mellett meg kell őrizni az eredeti operációs rendszert és azokat a fájlokat, amelyek az alkalmazói szoftver működéséhez szükségesek. Emellett le kell írni az ún. emulátor-specifikációkat, azaz az eredeti hardverplatform minden olyan attribútumát (például sebesség, display-attribútumok, eszközök és perifériák stb.), ami a jövőben is lehetővé teszi az aktuális emulátor programozását.<sup>68</sup> Az emulálással – a hívei szerint, akik többnyire informatikai szakemberek – kiküszöbölhetők a migrálási stratégia hiányosságai, mint pl. hogy jelentősen megváltozik az iratok formája, csorbul a hitelesség és a használhatóság, a konszekutív konvertálások pedig hosszú távon halmozott adatvesztést eredményeznek.

Mindez igaz, a levéltárak azonban mind a mai napig vonakodnak az emulálástól. A migrálás előbb sorolt hiányosságait a levéltárosok nem érzik olyan súlyosnak, ezért definiálták az irat lényegét, a hitelességet szerintük nem a változatlan forma vagy bitsorozat biztosítja (5. és 6. fejezet), a konszekutív konvertálás pedig elkerülhető. Az emulálással elérhető teljes értékű használhatóság sem egyértelmű előny, mert pl. a szervek elektronikus iratkezelő rendszerében egy sor olyan funkció is van (pl. új adat felvétele, meglévő adat módosítása, törlése, munkafolyamathoz kapcsolódó lekérdezések stb.), ami nem releváns vagy egyenesen nemkívánatos a levéltárban. David Bearman megállapítása szerint

---

<sup>68</sup> Az emulálásról jó áttekintést ad (bár erősen elfogult) Jeff Rothenberg: *Avoiding Technological Quicksand: Finding a Viable Foundation for Digital Preservation* (Council on Library and Information Resources, Washington 1999). Elérhető: <http://www.clir.org/pubs/reports/rothenberg/contents.html>

nem az informatikai rendszer funkcióit kell megőrizni, hanem az ügyiratot, nem szabad összetéveszteni az információt az ügyirattal.<sup>69</sup>

Az emulálásnak egyéb hátrányai is vannak:

- A sok alkalmazói szoftver használatának megtanulása a mennyiség növekedésével hatványozott nehézséget okoz mind a levéltárosoknak, mind a kutatóknak. Pár évtized alatt egy nagy levéltárban sok száz ilyen halmozódhat fel.
- A hardver és a szoftverkörnyezet folyamatos változásával egyre több emulátort kell létrehozni, azokat pedig – az aktuális levéltári e-környezet változásával – újra és újra meg kell újítani.
- A jogvédett alkalmazói szoftverek gyakran licenccéltárak, ami halmozódás esetén rendkívül sok pénz rendszeres kiadását eredményezi.
- A sok eredeti szoftver által kezelt sok elkülönülő adatállomány nem teszi lehetővé vagy nagyon megnehezíti az egységes levéltári rendszer kialakítását, az adatok egységes visszakeresését és nagyon komplikált az internetes adatelérés.

Mindezek ellenére az emulálást nem lehet teljesen elvetni a levéltárakban, mert az elektronikus iratok egy részénél lehetnek olyan fontos funkciók, amelyek migrálással nem, csak emulálással őrizhetők meg. Több sikeres emulálási projekt is megvalósult, ilyen pl. a PLANETS európai projekt keretében a Freiburgi Egyetem vezetésével kifejlesztett Global Access to Emulation Services (GRATE) vagy a holland vezetés alatt x86-os PC-k emulálására kifejlesztett Dioscuri emulátor,<sup>70</sup> vagy a szintén európai finanszírozású KEEP projekt (Keeping Emulation Environments Portable) eszközei és eredményei.<sup>71</sup>

### 9.3 Analóg és hibrid megőrzés

Abból kiindulva, hogy a hagyományos és az elektronikus irat között nincs lényeges különbség, a levéltárak kb. az 1990-es évekig gyakran kinyomtatták a fontosabb e-iratokat, és így őrizték azokat, az e-verzióval viszont nem sokat törődtek. Az adatbázisok, különösen pedig a hálózatok terjedésével azonban erről a módszerről gyorsan kiderült, hogy teljesen alkalmatlan az elektronikus iratok teljességének és bonyolult belső és külső struktú-

---

<sup>69</sup> Idézi Stewart Granger: *Emulation as a Digital Preservation Strategy* (D-Lib Magazine, October 2000, vol. 6. No. 10.). Elérhető: <http://www.dlib.org/dlib/october00/granger/10granger.html>

<sup>70</sup> Elérhető: <http://www.planets-project.eu/>, [http://www.planets-project.eu/docs/reports/Planets\\_PA5-D7\\_GRATE.pdf](http://www.planets-project.eu/docs/reports/Planets_PA5-D7_GRATE.pdf), <http://dioscuri.sourceforge.net/>

<sup>71</sup> Elérhető: <http://www.keep-project.eu/ezpub2/index.php>

rájának reprodukálására. Egy kinyomtatott e-mail nemcsak a funkcionalitását veszíti így el, hanem elvesznek a beágyazott fájlok, csatolmányok, egy kinyomtatott weblapnál nem tudunk mit kezdeni a még a lapon belüli linkekkel sem, egy adatbázis kinyomtatása pedig egyszerűen értelmetlen.

Az ezredforduló után megfordult az irány és manapság már inkább az a jellemző, hogy a szervezetekhez, vállalatokhoz beérkező papíriratokat digitalizálják, majd az eredetit kiselejtezik, így hozva létre egy egységes e-rendszert. Segítette ezt a trendet az is, hogy a fejlett országokban a jogalkotás tisztázta az e-iratok hitelességének feltételeit, és kiépült az ehhez szükséges infrastruktúra (pl. az digitális aláírások kibocsátásához). Egyes kormányzatok kifejezetten bátorítják a keletkező papíriratok mennyiségének csökkentését, az ausztrál kormányzat pl. 2011-ben ezzel a céllal elindította a Government's Digital Transition Policy-t, az indoklás szerint takarékosági okok miatt.<sup>72</sup>

Mindazonáltal vannak olyan iratok, amelyek papírformában való megőrzése kívánatos vagy egyenesen jogszabály írja elő. Magyarországon pl. törvény írja elő, hogy a parlament által hozott törvények hivatalos, eredeti példányait a Magyar Nemzeti Levéltárnak kell őriznie. Szakmai szempontból nézve olyan újonnan keletkezett irategyütteseket érdemes papírformában is őrizni, amelyek kiemelt jelentőségűek, az eredeti formájuk és hitelesítésük (pl. aláírás) megőrzése is nagyon fontos, valamint a külső struktúrájuk egyszerű. (Ilyenek tipikusan az egyszerű sorozatok, de az egymásba kapcsolódó ügyiratok már nem.)

Az elmúlt évtizedek tapasztalatai alapján szinte biztosra vehetjük, hogy a könyvek és iratok területén belátható ideig még párhuzamosan létezni fog a papír- és az e-világ. Egyrészt azért, mert a papír – minden hátránya ellenére – még mindig természetesebb hordozónak és megjelenítőnek számít a képernyővel összehasonlítva. Az audio-anyagok és a (mozgó)képek tekintetében, ahol nem áll fenn ilyen különbség, már teljesen eltűntek az analóg technikák. Másrészt az iratképzők többsége nem engedheti meg magának vagy nem érzi annak kényszerét, hogy egy költséges és bonyolult e-iratkezelő rendszert használjon, olyant, ami kizárja vagy a minimálisra csökkenti a papíriratok számát.

A mikrofilmeknek is lehet szerepük az e-iratok hosszú távú megőrzésében. A mikrofilmnek három nagy előnye van: 1) nagyon sűrű információt tartalmaz, 2) nagyon stabil, ha megfelelő körülmények között őrzik, akkor több száz évig is eltart, 3) relatíve hardver- és szoftverfüggetlen (az írás-olvasás szempontjából).

---

<sup>72</sup> Lásd: <http://www.naa.gov.au/records-management/digital-transition-policy/scanning-incoming-paper.aspx>

Az 1980-as években népszerű volt a Computer Output Microfilm (COM) technológia, aminek az a lényege, hogy a számítógépes kódokat egy speciális berendezéssel mikrofilmre fényképezik. Az elképzelés az volt, hogy analóg módon, mikrofilmen lehet tárolni a számítógépes kódokat, amiket később, ha szükség van rá, digitalizálni lehet. Bár a COM-technológia az utóbbi évtizedekben eléggé visszaszorult, leírni még korai lenne, mert az előbb felsorolt előnyök továbbra is fennállnak. A svéd Nemzeti Levéltár pár éve vezette be a COM-ot az e-iratok tárolására, és számításuk szerint a mikrofilm tárolási költsége csak töredéke a digitális tárolásnak.<sup>73</sup>

De szerepet kaphat a mikrofilm a papíriratok digitalizálásában is. A német Bundesarchiv stratégiája szerint a digitális másolat főként kommunikációs és állományvédelmi funkciót tölt be (azaz ezt használják kutatásra és ezáltal kímélik a papíriratokat), a párhuzamosan készülő vagy már meglévő mikrofilm-másolat pedig biztonsági másolatként szolgál (azaz, ha a papírirat elpusztul, akkor ez lép a helyére).<sup>74</sup>

#### 9.4 Különleges iratok

Vannak olyan e-iratok, amelyeknek vagy nincs megfelelőjük az analóg világban, vagy ha igen, akkor jelentősen különböznek tőlük. Közülük hármat szeretnék kiemelni, mert jelentőségük és különlegességük miatt külön figyelmet érdemelnek a megőrzési stratégia kidolgozásánál.

##### 9.4.1 *E-mail-ek*

Az e-mail valószínűleg a legnagyobb tömegben keletkező e-irat, a Radicati Groupe becslése szerint 2011-ben az e-mailfiókok száma kb. 3,1 milliárd volt, egy átlagos üzleti felhasználó átlag naponta 33 üzenetet küldött el (az ún. levélszemeteket [spam] nem számították).<sup>75</sup> Az e-mailek már nemcsak a hagyományos levelek, hanem az adminisztratív feljegyzések helyét is átvették, megjelentek minden döntési szinten és minden adminisztratív folyamatban. Ehhez képest a szervek csekély figyelmet fordítanak rájuk, pedig hogy milyen fontos információkat tartalmaznak, azt jól bizonyítják az elmúlt évek nagy nyugati botrányai és pereit, mint pl. az Enron-, az Arthur Andersen-, a Wikileaks- vagy az Apple contra

---

<sup>73</sup> Heather Brown et al.: *Instalment on "The role of microfilm in digital preservation"* (DCC Digital Curation Manual 2011) 15.old. Elérhető: [http://www.dcc.ac.uk/sites/default/files/documents/Microfilm\\_2011\\_Final.pdf](http://www.dcc.ac.uk/sites/default/files/documents/Microfilm_2011_Final.pdf)

<sup>74</sup> *Digitalisierung im Bundesarchiv 2011-2016*. Elérhető: [http://www.bundesarchiv.de/imperia/md/content/bundesarchiv\\_de/fachinformation/informationstechnologie/digitalisiertesarchivgut/digitalisierungimbarch2011\\_2016fv.pdf](http://www.bundesarchiv.de/imperia/md/content/bundesarchiv_de/fachinformation/informationstechnologie/digitalisiertesarchivgut/digitalisierungimbarch2011_2016fv.pdf)

<sup>75</sup> *The Radicati Group Inc. Email Statistics Report, 2011-2016*, 2-3. old. Elérhető: <http://www.radicati.com/wp/wp-content/uploads/2011/05/Email-Statistics-Report-2011-2015-Executive-Summary.pdf>

Samsung-ügy, amelyekben döntő szerepet játszottak a kiszivárgott vagy feltárt e-mailek.<sup>76</sup> Sajnos a levéltárak gyakran hiányos vagy rosszul rendezett e-maileket kapnak a szervektől és pedig több ok miatt.

A hivatalban keletkezett vagy kapott e-mail elméletileg hivatalos irat, ezért úgy kellene tekinteni és kezelni, mint egy eljárás produktumát, ami az adott eljárás többi iratához tartozik és az irattár szerves részét alkotja. A gyakorlat sokszor mást mutat. Korábban említettem, hogy az e-kommunikáció, benne az e-mail sok tekintetben átvette a szóbeli kommunikáció funkcióját és jegyeit (pl. az írásbeli szabályok, a formai jegyek eltűnnek vagy elmosódnak) – ez kikezdi az irat belső struktúráját, pl. tárgymegjelölés híján megnehezíti a besorolását. A szerveknél a dolgozók általában a nevük alapján kapnak e-mail-fiókot és sokan ezt használják a privát levelezésükre is. Az eredmény: a hivatalos és a magánlevelek keverednek, ami nagyon megnehezíti az archiválást, különösen akkor, ha a jogszabályok erősen védik a magánéletet.<sup>77</sup> A mailek tárolódhatnak a hivatal központi szerverén vagy a dolgozó számítógépén, ez utóbbi esetben nem ritka történet, hogy az intézményből távozó dolgozó kiüríti a teljes e-mail-fiókját. Mindezek összességben azt eredményezik, hogy az intézményi e-maileket sokan inkább magániratnak tekintik.

Persze vannak olyan megoldások, amelyek erősítik a hivatalokban keletkezett e-mailek hivatalos jellegét és hitelességét, ilyen pl. ha az e-mailezés központi rendszerben történik és bizonyos (meta)adatokat kötelező kitölteni (különben a rendszer nem engedi a mailt elküldeni), ha fióknév nem a dolgozó nevére, hanem a funkciójára utal (pl. osztalyvezeto-x-ugyosztaly@y-miniszterium...), amit az utódja örököl, illetve ha nem a dolgozó gépén tárolják az üzeneteket.

Az e-maileket gyakran külön kezelik, főként azért, mert külön e-mail rendszert működtetnek, így viszont elvágják azokat a kapcsolatokat, amelyek az iratokat összefűzik. Amíg él az ügy, a dolgozók tudják, hogy mi tartozik össze, de az idő múlásával egyre nehezebb rekonstruálni egy-egy ügyiratot.

Természetesen erre is vannak megoldások. A legjobb az, ha az e-mail rendszer a szerv E-Records Management Systemjének (ERMS) a része, így az üzenetek mindig a megfelelő helyre kerülnek. Ha a szervnek nincs ERMS-e, akkor az e-mail rendszert kell úgy működtetni, hogy minden e-mail megkapja az irattári azonosítóját, ami összeköti az

---

<sup>76</sup> Néhányat idéz Christopher J. Prom: *Preserving email* (DPC Technology Watch Report 11-01, 2011), 5.old. Elérhető: <http://www.dpconline.org/advice/technology-watch-reports>

<sup>77</sup> Ez természetesen országonként változik, Franciaországban pl. nagyon erős ez a védelem. Lásd Frédéric Brégier et al.: *L'archivage des messageries électroniques, Preuve de concept VITAM* (Version 1.1, 2013) 15-16.old.

ügy többi iratával, és olyan mappastruktúrát kell kialakítani, ami szinkronban van az irattári besorolással.

A csatolmányok, amelyek gyakran fontosabbak, mint maga az üzenet, komoly gondot okoznak az e-mailek kezelésénél, de különösen az archiválásnál, főként mert rendkívül változatos formátumúak. Ha ERMS-be vagy archiváló rendszerbe kerülnek, érdemes őket külön őrizni az e-mailtől – de természetesen az összetartozást metaadatokkal fenntartani –, esetleg szabványos formátumúakra konvertálni.

A levéltárba szállítás után az e-mailek sorsa ugyanaz lesz, mint a többi e-iraté: ellenőrizni, szabványos formátumokra konvertálni kell (többnyire az XML-formátumot ajánlja a szakirodalom<sup>78</sup> az emulálás viszont nem javallott), le kell írni és be kell illeszteni őket a levéltár egységes nyilvántartó és visszakereső rendszerébe.

#### 9.4.2 Adatbázisok

A Nemzetközi Levéltári Tanács Elektronikus iratokra vonatkozó kézikönyve a következő definíciót írja az adatbázisról: „... összekapcsolt, szervezett (strukturált) adatok együttese. Két lényeges részt foglal magában: egyedi adatokat vagy elemeket (tartalom), valamint azon struktúrákat, amelyek szerint ezek az adatok szervezettek.” Majd később ezt írja: „A modern adatbázisok nem korlátozódnak relációs szerkezetre, mivel azokba szövegfájlokat, képeket és hangdokumentumokat stb. is integrálnak.”<sup>79</sup>

Valószínűleg az adatbázisok a legnehezebben archiválható e-iratok és pedig több okból kifolyólag, amelyeket már érintettem az 5. fejezetben:

- nagyon bonyolult a struktúrájuk;
- a működés során átmeneti (tranzakciós) rekordokat is létrehozunk attól függően, hogy hogyan csoportosítjuk (kérdézzük le) az adatokat;
- a legtöbb adatbázis dinamikus, azaz a tartalmuk gyorsan változik;
- a hosszú ideig működő adatbázisok szerkezetben és funkcióban is változnak.

A megoldás csak kompromisszumos lehet: a struktúrából és az adatok szervezéséből (ezt nevezik adatbázis sémának) annyit kell átmenteni az archiváláshoz, amennyit lehet, az

---

<sup>78</sup> Lásd pl. Frédéric Brégier et al. i.m. 36-37. és 48-52. old., illetve Christopher J. Prom i.m. 23-24. old., illetve Maureen Pennock: *Instalment on "Curating E-Mails: A life-cycle approach to the management and preservation of e-mail messages"* (DCC Digital Curation Manual 2006) 48-52.old. Eelérhető: <http://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/curating-e-mails/curating-e-mails.pdf>. Ezekben a művekben a szerzők áttekintik a tárggyal kapcsolatos nemzetközi kutatásokat.

<sup>79</sup> *Guide pour la gestion archivistique des documents électroniques* (Conseil international des archives, Études 8. 1997) 43-44.old.

adatokról pedig időről-időre inkrementális beszállítást kell kérni (csak a megváltozott rekordokat tartalmazza), és a struktúra vagy a funkció változásait is rögzíteni kell.

Az adatbázisoknál csak kivételes esetekben javallott az emulálás, akkor ha lényeges funkciói csak így őrizhetők meg. A 9.2. fejezetben leírt emulálás elleni érvek – az alkalmazói szoftverek felhalmozódása, licenstdíjak, elszigetelt adatbázisok – itt is érvényesek. Emellett nem ritka eset, hogy a szerv és a szállító közötti szerződés megtiltja a jogvédett szoftver átadását harmadik személynek, így a levéltárnak is.<sup>80</sup> Viszont a migrálási stratégiát választó levéltárak sem tudnak minden adatbázishoz külön kezelő szoftvert készíteni, ezért többre közös, egyszerűsített kereső- és kezelőprogramokat alkalmaznak. Ezen kívül mindenképpen kívánatos egy egységes, az egész állományra érvényes kereső rendszert is kiépíteni.

Talán a probléma bonyolultsága miatt csak kevesen próbáltak univerzális megoldást kidolgozni az adatbázisok archiválására. Az egyik ilyen Svájci Szövetségi Levéltár SIARD-ja (Software Independent Archiving of Relational Databases)<sup>81</sup>, egy megőrzési szoftver relációs adatbázisok tartalmának archiválására. Természetesen nemcsak a tartalmat, hanem a metaadatokat és a kapcsolatokat (relation) is átmentik, így lehetővé válik egy új adatbázis építése. Kifejlesztettek SIARD-nak nevezett specifikus formátumot, ami nyílt szabványokra épül: XML, SQL, ZIP et UNICODE. Ellentétben egy ún. „lapos” adatbázis-exporttal, a SIARD lehetővé teszi az adatbázis struktúrájának kódolását (táblalisták, minden tábla mezőinek és minden mezőtípus leírása). A formátum azt is lehetővé teszi, hogy a táblák közötti kapcsolatokat is leírják, valamint a kezelési információikat.<sup>82</sup>

A Svájci Szövetségi Levéltár elérhetővé teszi a szoftvert az internetes honlapján keresztül.

A Bajorországban kifejlesztett CHRONOS nevű rendszer fő célkitűzései hasonlóak, mint a svájci társáé.<sup>83</sup>

- leválasztani az adatokat az eredeti adatbázis-kezelőtől;
- megőrizni az eredeti struktúrát és az adatok szemantikáját szabad szemmel olvasható formában;

---

<sup>80</sup> Megjegyzendő, hogy néha maga az adatbázis is jogvédett, ami a konvertálásnak lehet az akadálya.

<sup>81</sup> Elérhető: <http://www.bar.admin.ch/dienstleistungen/00823/00825/index.html?lang=fr>

<sup>82</sup> <http://www.archivesdefrance.culture.gouv.fr/static/4163>, 2. old.

<sup>83</sup> Stefan Brandl, Dr. Peter Keller-Marxer: *Long-term Archiving of Relational Databases with Chronos* (First International Workshop on Database Preservation, Edinburgh 2007). Elérhető: [http://homepages.inf.ed.ac.uk/hmueller/presdb07/papers/Germany\\_CHRONOS\\_PresDB07.pdf](http://homepages.inf.ed.ac.uk/hmueller/presdb07/papers/Germany_CHRONOS_PresDB07.pdf)

- a sok adatbázist egy levéltári adatbázisba integrálni;
- inkrementális módon gyarapítani az adatokat;
- online, web-alapú hozzáférést biztosítani;
- szöveg alapú export-import funkciókat kiépíteni.

Érdemes kiemelni, hogy a Chronos tiszta szövegformátumú levéltárakat (adatbázisokat) produkál (ASCII/UTF-t az elsődleges adatoknak és XML-t a metaadatoknak). Ez fontos biztonsági tényező, mert hardver- és szoftverfüggetlenné teszi az adatokat, és ennek eredményeként akár egy szimpla webböngészővel is el lehet elérni azokat.

### 9.4.3 Weblapok

A weblapokat mind a könyvtárak, mind a levéltárak gyűjtik – a könyvtárak jóval előrébb tartanak –, és bár vannak párhuzamosságok, azért a szempontrendszerük általában eltérő. A könyvtárak kiadványként tekintenek a weblapokra, mint a publikált kultúra lenyomatára, míg a levéltárak az iratképzők (intézmény vagy személy) fondjának részére.<sup>84</sup> Ennek megfelelően a levéltárak gyűjtési politikája általában sokkal célzottabb, pl. míg a British Library az Egyesült Királyság web-archív-ját építi, azaz a célja a teljes társadalom dokumentálása, addig a londoni Nemzeti Levéltár Web Continuity Project célja „a kormányzati web-anyag átfogó archiválása”.

Bár az intézményi honlapokat joggal tekinthetjük az illető szervezetek fondjai részének, azért van különbség a rendes adminisztrációban képzett iratok és a világhálóra kitett információk között. A weben elérhető intézményi információnak mindig van egy publikus része – az, amit a szervezet meg akar mutatni magából –, ott általában naprakész információk vannak jól strukturált formában. Egyre több anyag és intézményi tevékenység kerül ki a honlapokra, így pl. digitális kiadványok, kiállítások, kutatói adatbázisok, digitalizált iratok. A publikus rész mellett sok intézmény működtet egy olyan szakmai vagy közösségi fórumot, ami összefogja az adott közösséget, illetve annak információit, és tájékozási pontként szolgál. Ez a kommunikációs felület már gyakran zárt, csak a tagok számára van fenntartva, aminek komoly következménye van az archiválásra.

A weblapok archiválását jelentősen megnehezíti a bonyolult szerkezetük és az instabilitásuk, ez azonban másféle bonyolultság és változékonyság, mint amit az adatbázisoknál

---

<sup>84</sup> Persze vannak levéltárak – különösen Kanadában és Hollandiában –, amelyek a total archives (teljes levéltár) koncepcióját gyakorolják, azaz nemcsak szervek, magánszemélyek fondjait gyűjtik, hanem a kompetenciájukba eső terület (ország, megye, témakör stb.) összes emlékét, legyen az irat, könyv, fénykép stb.

tapasztalhatunk. A weblapok sok beágyazott iratot, sokféle formátumot és rengeteg, más weblapra mutató linket tartalmaznak, ez utóbbiak működőképes megőrzése – ahogy azt az 5. fejezetben már említettem – teljesen reménytelen, mert ha követnénk azokat, akkor az archiválendő anyag mennyisége hatványozottan növekedne.

A weblapok dinamikus jellege elképesztő mértékű. Számosan próbálták számszerűsíteni a változékonyságot, álljon itt néhány adat: egy 1999. évi felmérés szerint az egy évvel korábban létrejött web-oldalak 44%-a már nem létezett vagy jelentősen módosították.<sup>85</sup> Más kutatások a weblapok átlagos élettartalmát 44 és 100 nap közötti időtartamra becsülték. Az online tudományos publikációkban a hivatkozások, idézetek közel harmada 10 év után ellenőrizhetetlen volt, ami kikezdi a tudományosságot.<sup>86</sup> Örökös dilemma, hogy milyen gyakorisággal készüljenek „pillanatfelvételek” a változékony weblapokról, ezt a lap fontossága, a különleges események vagy a tartalom gyors változásának függvényében kell meghatározni. A korábbiakban többször említettem a levéltári e-iratok mennyiségének gyors növekedését, mint alapvető problémát, nos ezért nem kis részben a weblapok „pillanatfelvételei” felelősek, ráadásul nagy bennük a redundancia: egyes felmérések szerint 25 százalékban duplikált vagy többszörösen ismételt információt tartalmaznak.<sup>87</sup>

Nagyon sok a jelszóval védett tartalom, számos weblap mögött pedig nagy adatbázis van, onnan lehet lekérdezni az adatokat. Ezeket a „rejtett” tartalmakat az általános gyűjtést végző könyvtárak vagy levéltárak nem tudják elérni. A gyűjtési technikai módszer kétféle lehet: kliens oldali és szerver oldali. A kliens oldali archiválás – ez a gyakoribb, automatikusan működő ún. web-crawler-ekkel végzik – azt jelenti, hogy „kívülről” lemásolják a weblap tartalmát. Ezzel viszont nem lehet archiválni a weblapon működő adatbázisokat és a jelszóval védett tartalmakat, ez csak úgy lehetséges, hogy a másolás közvetlenül a szerverről történik, amihez viszont a működtető együttműködése szükséges.<sup>88</sup>

A weblap archiválók könnyen beleütközhetnek a szerzői jogokba. Egyenként beszerezni az engedélyt a lap jogainak birtokosától nehézkes, nagyszámú gyűjtésnél egyenesen lehetetlen. A nagy archiválók – mint pl. az Internet Archive, amely a világ minden részéről

---

<sup>85</sup> Peter Lyman: *Archiving the World Wide Web* (Building a National Strategy for Digital Preservation: Issues in Digital Media Archiving; Council on Library and Information Resources and Library of Congress; Washington D.C. 2002) 38. old.

<sup>86</sup> Maureen Pennock: *Web-Archiving* (DPC Technology Watch Report 13-01, 2013, Digital Preservation Coalition) 3-4. old. Elérhető: <http://www.dpconline.org/advice/technology-watch-reports>

<sup>87</sup> Daniel Gomes & André L. Santos & Mário J. Silva: *Managing Duplicates in a Web Archive*, 4. old. Elérhető: <http://visibilidade.net/daniel/docs/presentations/gomes06duplicatesPPT.pdf>

<sup>88</sup> Maureen Pennock, i.m. 7. old.

végez gyűjtést<sup>89</sup> – a kizárásos opciót (opt out) alkalmazzák, azaz akkor törlik a tartalmat, ha szerzői jogok birtokosa ezt külön kéri. A nemzeti könyvtárak és levéltárak viszont általában törvényi felhatalmazás alapján dolgoznak.

A weblapok kinézete, formája fontosabb, mint az átlagos e-iratoknak, ezért arra kell törekedni, hogy ezt is archiváljuk. (Az általánosan használt formátum a HTML és az XML.) Gondot jelent viszont, hogy a kinézet erősen függ a böngészőtől, mivel azonban nem lehet mindenhová optimalizált böngészőt rendelni, kompromisszumos megoldást kell találni.

## 10 Az elektronikus iratok fizikai hordozóinak megőrzése

Bár az előzőekben a szakmai, technikai megőrzés fontosságát hangsúlyoztuk, ez nem jelenti azt, hogy a fizikai megőrzést, azaz az adatokat, információt hordozó médiát ne kellene fizikailag is megőrizni.

A levéltárakban hosszú távú megőrzésre használt adathordozókat két nagy csoportra lehet osztani: mágnesesekre és optikaiakra. Az első csoportba tartozóknál az adatokat mágnesezett (merev)lemez vagy szalag tartalmazza, a másodiknál (CD, DVD) a rögzítés, illetve az olvasás egyaránt lézerténnel történik.<sup>90</sup> A hordozók kiválasztásánál a következő szempontokat kell mérlegelni: stabilitás, élettartam, elterjedtség (szabványosság), kapacitás, kezelhetőség, adatmódosíthatóság (a hitelesség miatt fontos). Általánosságban elmondható, hogy a mágneses szalagok-lemezek kevésbé stabilak (hajlamosak a lemágnesződésre), az élettartamuk ezért rövidebb, mint az optikai hordozóké, eléggé nehezen kezelhetők (az adatokhoz való hozzáférés lassú, kivéve a merev lemezeket), az adataik módosíthatók, viszont széles körben használatosak, és nagy a kapacitásuk. Az optikai lemezek stabilak és hosszabb élettartamúak (igaz, mechanikus hatásra nagyon sérülékenyek), általánosan elterjedtek, a kapacitásuk viszont elég kicsi (de gyorsan nő), könnyen kezelhetők, az adataik pedig nem módosíthatók (kivéve az újraírható CD-eket és DVD-eket).

Az elektronikus adathordozók terén is rendkívül gyors a változás és a fejlődés, a szabványok 10-15 év alatt elavulnak, ami azt jelenti, hogy a ma használatos adathordozók 20 év múlva jó eséllyel már nem lesznek olvashatók. Ezért inkább elméleti jelentőségű egyes hordozók (jellemzően: CD-k és DVD-k) gyártóinak állítása, miszerint a termékük élettartama legalább 100 év. Ugyanis a hardver- és szoftverelavulási periódus biztosan rövidebb,

---

<sup>89</sup> 2013-as adat: több mint 150 milliárd weblapot archivál. Maureen Pennock, i.m. 5.old.

<sup>90</sup> A Pen drive-ok használata tárolásra, bármennyire is vonzó a nagy befogadóképességük, nem javallott.

mint ez az élettartam, azzal pedig nem megyünk semmire, ha van egy tartós hordozónk, viszont már nem áll a rendelkezésünkre az azt olvasni és kezelni tudó hardver és szoftver. Ez persze nem jelenti azt, hogy bármilyen médiát használhatunk, a rossz minőségű ugyanis pár év alatt is tönkre mehet. 2005-ben a francia levéltári vezetés egy általános vizsgálatot folytatott az ottani megyei levéltárakban őrzött CD-kkel kapcsolatban. A szűrő-próbaszerű tesztelés eredményei azt mutatták, hogy a CD-k több mint 20%-a rossz állapotban volt.<sup>91</sup>

A hordozó kapacitása és kezelhetősége szorosan összefügg. 10 terabájtnyi információ 13000–16000 CD-re (650-850 MB/db) vagy 1200-2200 DVD-re (4,7-8,5 GB/db) vagy 14-29 DAT/DLT szalagra (35-72 GB/db) írható. Mivel az állományt szisztematikusan figyelni kell, azaz időről időre minden hordozót tesztelni kell, hogy olvashatók-e még az adatok, nem mindegy, hogy ezt a műveletet sok ezer CD-vel vagy csak néhány tucat szalaggal kell-e végrehajtani.

Az állományvédelem hardver- és szoftverigényéről sokan elfelejtkeznek, pedig az előbb említett rendszeres ellenőrzés ezt megköveteli. Ha sok CD-vel vagy DVD-vel dolgozunk, akkor ehhez megfelelő berendezés szükséges (például jukebox). Külön szoftverek (repository management software) kaphatók állományvédelmi ellenőrzésre és kezelésre. Az ilyen szoftverek automatikusan elvégzik az adatok azonosítását, az integritás ellenőrzését, és az előírt időben elvégzik a frissítést és az utasításoknak megfelelő migrálást.<sup>92</sup> Nagyon fontos előre, szabályzatban rögzíteni, hogy milyen ellenőrzéseket milyen rendszerességgel kell elvégezni, a hibáknak milyen mennyiségénél és minőségénél milyen beavatkozásokat kell tenni. A hibamérésre vonatkozóan vannak nemzetközi szabványok és ajánlások, ezeket érdemes figyelembe venni.<sup>93</sup>

Mindenképpen ajánlott több példányban őrizni az elektronikus iratokat. Jó, ha van egy biztonsági példányunk, amit távoli helyen őrzünk, egy mesterpéldányunk, amit optimális raktári körülmények között tartunk, valamint egy munkapéldányunk. Akár mind a három lehet eltérő hordozón, például a biztonsági példány szalagon, a mesterpéldány DVD-n, a munkapéldány pedig merevlemezen.

---

<sup>91</sup> Lásd: <http://www.archivesdefrance.culture.gouv.fr/static/1053>

<sup>92</sup> Kevin Bradely: *Risks Associated with the Use of Recordable CDs and DVDs as Reliable Storage Media in Archival Collections – Strategies and Alternatives* (UNESCO, Paris 2006) 21. old.

<sup>93</sup> Erre jó útmutatást ad Kevin Bradely, i.m. 15. old.

A mágneses szalagok használatát minimalizálni kell. A szalagokban felhalmozódó feszültségek és az átmágneseződés csökkentése érdekében időnként (1-2 évenként) át kell tekercselni a kazettákat, és az egyik végükre csévélve kell tárolni őket. Nagy figyelmet kell fordítani a berendezések rendszeres karbantartására és tisztítására, mert egy nem megfelelő műszaki állapotú gép jóvátehetetlenül károsíthatja az adathordozót.

Mágneses adathordozókat nem szabad tápegységek, kábelek és egyéb mágneses teret generáló eszköz közelében tárolni, mert az információvesztést okozhat. Ennek mértéke a térerősségtől, a távolságtól és a hatás időtartamától függ, ezért a legkisebb mágneses tér is káros lehet, ha sokáig fejt ki a hatását. A szalagokat és a lemezeket az élükre állítva, függőlegesen kell tárolni.

A magas hőmérséklet és relatív páratartalom, illetve ezek ingadozása kerülendő, mert az alapanyag törékennyé válását, a mágneses réteg leválását vagy a szalag összeragadását okozhatják. A magas hőmérséklet a mágnesesség megszűnését, vagyis az információ elvesztését is eredményezheti. A 40% fölötti páratartalom gyorsítja a kötőanyag bomlását. A biztonsági, a mester- és a használati másolatok tároló helyének klimatikus követelményei eltérőek. A rövid, felhasználási célú tárolásra a normál szobai körülmények (20-24 °C, de mindig 55% relatív légnedvesség mellett) megfelelőek. Ezzel szemben a tartós megőrzés klimatikus feltételei sokkal szigorúbbak. A mágneses adathordozók számára ajánlott tárolási hőmérséklet  $18\pm 1^\circ\text{C}$ , a relatív légnedvesség  $45\pm 5\%$ .<sup>94</sup> Ha a tárolás és a használat helyének hőmérséklete között  $8^\circ\text{C}$ -nál nagyobb a különbség, akkor a különbségtől függően több órás akklimatizációt (köztes hőmérsékletű tárolást) kell beiktatni.

A városi levegőben előforduló korrozív gázok már kis mennyiségben is károsítják a mágneses szalagokat, ezért nagy hatékonyságú szűrők alkalmazása válhat szükségessé a raktárakban. A mágneses felvételeket fénytől, különösen napfénytől védve kell tartani. A tároló- és használati hely tisztaságára fokozott figyelmet kell fordítani, mert a mágneses felületen a por, zsiradék, vegyi szennyeződések páralecsapódást és káros vegyi folyamatokat okozhatnak.

Az optikai adathordozókat mechanikai károsítás szempontjából leginkább a hajlítás és karcolás veszélyezteti, ezért az ilyen hatásokat kerülni kell. A lemezek kézbevételekor szálmentes kesztyű viselése ajánlott, mert az ujjlenyomat olvashatatlanná teheti a lemezt. Tisztításra nem szabad oldószereket használni. A por vagy enyhe szennyeződés textíliá-

---

<sup>94</sup> Meg kell mondani, hogy a különböző nemzeti szabványok eléggé eltérő paramétereket ajánlanak, különösen az RH tekintetében.

val vagy puha ecsettel eltávolítható. Itt is alapvető követelmény, hogy a lejátszó berendezés tiszta és rendszeresen karbantartott legyen. Nem szabad a lemezeket a lemezmeghajtóban hagyni.

Az optikai adathordozók tárolásának klimatikus követelményei hasonlóak, mint a mágneseseké. Hűvösebb tárolással a természetes öregedési folyamatok lassíthatók. Nagyon fontos a klimatikus tényezők stabilitása. A relatív páratartalom ingadozása, a meleg erősen felgyorsítja a károsodási folyamatokat. A 60% fölötti relatív páratartalom kedvez a penészek megtelepedésének. Megfelelő intenzitású légmozgással a raktárban a kedvezőtlen mikroklímák kialakulása megakadályozható.

Közvetlen napfény hosszabb ideig nem érheti a lemezeket. Nem szabad a lemezeket egymásnak szoruló halmokban tárolni, mert könnyen deformálódhatnak. Az eredeti (műanyag) dobozok a deformálódás ellen is védelmet nyújtanak.

## **11 A levéltárban átalakulása az elektronikus iratok korában**

Az e-világ átalakította a levéltárat, mert amint az előző fejezetekben láttuk, maguk az iratok is átalakultak: a struktúrájuk bonyolultabb lett, az elemeik elváltak és súlyuk megváltozott, a metaadatok kiemelt szerepet kaptak, az iratok technikai jellege előtérbe került, a szabványosítás elemi szükségletté vált. Az e-iratok integritása és hitelessége biztosításának nehézségeit már korábban részleteztem (5-6. fejezetek), most két olyan tényezőt szeretnék megvilágítani, amelyek szintén erősen befolyásolták a levéltári módszereket: az egyik a harmadik dimenzió elvesztése, a másik pedig a mennyiségi robbanás.

Az e-iratok, ha megjelenítik őket, csak két dimenzióban léteznek, ez viszont azt eredményezi, hogy rendkívül nehéz bennük a tájékozódás. Ha a számítógép egyik mappájában látunk több ezer rendezetlen fájlt, akkor az állomány gyakorlatilag elveszettnek tekinthető, ha ugyanezzel a mennyiséggel papírirat formájában szembesülünk (néhány doboznyi irat), akkor nem reménytelen a rendezés. Az e-iratokban csak akkor tudunk rendet tartani, akkor tudunk bennük tájékozódni, keresni, ha darabszinten metaadatolva vannak, ami az analóg világban nem feltétlen követelmény.

Korábban említettem, hogy a levéltárosok évszázadokig darabszinten igyekeztek leírni az iratokat. Az akkori leírás, amit ma metaadatolásnak neveznénk, egyszerre szolgálta az iratokhoz való hozzáférést (hogy megtalálják a keresett adatokat) és a nyilvántartást (ami a hitelességet szolgálta), mert az azonosítóval ellátott iratok meglétét bizonyította, és a

helyét rögzítette az irattári rendszerben. Ilyen tipikus középkori és újkori leírások (segédletek, nyilvántartások) voltak a különböző katalógusok, listák és elenchesok. Az egyes iratok tudományos vizsgálatának szabályait az oklevéltan fogalmazta meg.

A írásbeliség fejlődése a megőrzendő iratok gyors mennyiségi növekedését eredményezte, és új nyilvántartási és visszakeresési rendszerek kifejlesztését tette szükségessé. Az egyre fejlettebb és növekvő egyházi és állami bürokrácia fokozatosan elhagyta a darabszintű leírást, és áttért a tárgyi vagy formai jegyek alapján kialakított iratcsoportok (sorozatok), majd később az ügyiratok (az egy ügygel kapcsolatos iratok együttesének) kialakítására és leírására. A 19. században a közigazgatás terjeszkedésével párhuzamosan tovább nőtt a képzett iratok mennyisége, és amint a 2.1. fejezetben említettem, a mesterséges tárgyi vagy kronológiai besorolások már nem voltak alkalmasak a megfelelő információ-visszakeresésre. Ezért dolgozták ki a levéltárosok a század végén a provenienciaelvet, aminek segítségével nagy biztonsággal lehetett kinyerni releváns információt nagytömegű iratból, és az iratok hitelessége is biztosítható volt. A provenienciaelvet követő levéltáros az iratok külső szerkezetéből, tartalmából és a levéltári kontextusokból eredő szintetikus információkat állapítja meg, ez alapján értékeli és ír le. Az origó mindig a fond, ami egy iratképző terméke, az ő tevékenysége, jogállása, funkciója a legfontosabb információ a leírásban. A provenienciaelven alapuló levéltártan fondtól lefelé, azaz az általánostól az egyediig elemel, és mindig szem előtt tartja irategyüttesek összetartozását, azaz intellektuális egységét. Ennek az elvnek a legegységelműbb megfogalmazása az ISAD(g) leírási szabványban olvasható. A levéltártan tehát – ellentétben a diplomatikával – nem az egyedi iratok, hanem az iratcsoportok (aggregációk) vizsgálatát részesíti előnyben.

A levéltárosok szemléletében történt némi „visszarendeződés” a 20. század közepén, amikor intenzív, tanácsadói és ellenőrzési kapcsolatot kezdtek kiépíteni az irattárakkal, a records manager-ekkel, így próbálván minőségi iratokat biztosítani maguknak a majdani beszállításkor. A records manager sokat foglalkozik egyedi iratokkal, és a szemlélethorizontja csak fondszintig terjed. Az iratok beszállítása után a levéltárosnak viszont az a legfőbb szakmai feladata, hogy a sok fondból kiépítsen egy új, átfogó rendszert, ezért számára az iratok külső struktúrája – az iratok, irategyüttesek közötti kapcsolatok – sokkal fontosabb, mint a belső.

Az e-iratokkal azonban változott a helyzet. A korábban (3. fejezet) ismertetett változások, mint az iratok szerkezetének módosulása, tartalmának, formájának és hordozójának

elválása, a technikai (informatikai) specifikációk megjelenése és minden elem sérülékenysége, illetve a sérülések gyakori rejtettsége újra előtérbe állította az egyedi iratok, a belső iratszerkezet vizsgálatát. Ugyanakkor bekövetkezett az e-iratok robbanásszerű mennyiségi növekedése is,<sup>95</sup> ami miatt viszont az egyedi iratok manuális vizsgálata teljesen reménytelen lenne, ezért olyan módszerekre (szabványokra) és eszközökre (szoftverre) volt szükség, amelyek lehetővé tették az e-iratok automatikus metaadatolását és integritásának ellenőrzését pl. beszállításkor vagy konvertáláskor, ellenőrzendő, hogy történt-e adatvesztés.<sup>96</sup> Az eddigi levéltári módszerek és technikák erre nem voltak alkalmasak, ehhez részben vissza kellett térni a diplomatika módszereihez, de még inkább az információtudomány segítségét kellett kérni. Az információtudomány legnagyobb „hozádéká” az, hogy általa oly módon lehet megszervezni az adatokat, hogy az eljárások automatizálhatóvá válnak. (Az információtudomány az egyedi irat vs. irataggregáció szempontjából semleges, mert mindkettővel foglalkozik.) A diplomatika hatása nyilvánvaló pl. az InterPARES projektben – ezt hangsúlyozza is<sup>97</sup> –, az információtudomány termékei pl. az InSPECT „lényeges jellegzetességei” (5. fejezet) vagy a teljes OAIS.<sup>98</sup>

A sok fond irataiból álló levéltári rendszer kiépítése azonban olyan szintetizáló feladatot is megkíván, amit sem a diplomatika, sem az irattan, sem az információtudomány nem tud elvégezni, vagy azért, mert nem a feladata, vagy mert nem képes erre. Ha pl. a levéltáros tanulmányozza egy szerv iratanyagán keresztül a működését, azt párhuzamba állítja az adott kor politikai vagy gazdasági viszonyaival, és ez alapján sorolja be a fondot a makrostruktúrába, akkor ilyen szintetizáló feladatot végez. Ugyancsak ezt csinálja, amikor a fond szervtörténeti ismertetőjében (kontextus) mindezt leírja, ezzel egy fontos – igaz, szubjektív – információval ellátva el a jövőbeni kutatót. Az informatika belátható ideig még nem képes olyan szintetizáló feladat elvégzésére, mint a levéltáros, mert rugalmatlan, nem képes úgy elvonatkoztatni, mint a kreatív ember. Ugyanakkor képes korlátozott absztrakciót elvégezni, pl. hatalmas adattengerből különböző szempontok alapján szűrni, előre programozott összefüggéseket kiemelni.

---

<sup>95</sup> A világon keletkezett (becsült) e-adatok számára vonatkozóan lásd pl. [http://www.planets-project.eu/docs/comms/PLANETS\\_BROCHURE.pdf](http://www.planets-project.eu/docs/comms/PLANETS_BROCHURE.pdf), 3. old. vagy a 9.4.1. fejezetben említett statisztikákat az e-mailekkel kapcsolatban.

<sup>96</sup> Az analóg világban erre nem volt szükség, csak az iratok külső struktúrájának a megőrzése kapott állandó hangsúlyt, a forma vagy a belső szerkezet megmaradása, ha az irat maga megmaradt, magától értődött.

<sup>97</sup> Lásd pl. [http://www.interpares.org/display\\_file.cfm?doc=ip1\\_atf\\_report.pdf](http://www.interpares.org/display_file.cfm?doc=ip1_atf_report.pdf)

<sup>98</sup> A levéltártan itt vázolt fejlődése ugyancsak leírható a hegeli hármas tétellel (tézis, antitézis, szintézis).

Az e-iratok világában a levéltárosnak adaptálnia kell az információtudomány vonatkozó módszereit és eszközeit, de ugyanakkor nem szabad szem elől tévesztenie az értékelő és leíró (szintetizáló) feladatát sem, különösen magasabb iratszinteken (fond, állag). Ha nem sikerül ezt az egyensúlyt megtartani, akkor sérül a levéltári munka teljessége, és ekkor beszélhetünk a levéltári és az irattári paradigma konfliktusáról.

A levéltári paradigma hívei szerint a levéltár a már nem aktív kormányzati és magániratok szervezett együttese, függetlenül attól, hogy bekerült-e már a levéltári intézménybe vagy sem. Ebben a felfogásban az iratok bizonyító és emlékezeti (tartalmi) funkciója egyensúlyban van, azaz egyaránt fontos mindkettő. Az irattári paradigma hívei szerint viszont az irat elsősorban bizonyíték az iratképzőnél és a vele kapcsolatban állóknál történt eljárásokról, ügyletekről, és amit főleg adminisztratív, illetve jogi okok miatt őriznek.<sup>99</sup> Ennek a felfogásnak a követői tulajdonképpen az irat funkcionalitását, bizonyító erejét, kontextusát, valamint formai-strukturális jegyeit abszolútizálják, a tartalmat, azaz az emlékezet-funkciót pedig degradálják. Az irat vizsgálatakor, értékelésekor igyekeznek maximálisan az objektív tényekre hagyatkozni (az előbb említett funkcionalitásra, jogszabályi előírásokra, formai-strukturális jegyekre), és így vélik megtalálni az iratra vonatkozó „abszolút igazságot” (mi az irat, mi az értékes és az értéktelen).<sup>100</sup> És – tegyük hozzá – az egyes iratokra fókuszál, az iratcsoportokat, a külső struktúrát többé vagy kevésbé negligálja. Ha felidézünk az InterPARES erős kötődését a diplomatikához, akkor láthatjuk, hogy felvázolt paradigma-konfliktus nem alaptalan.<sup>101</sup>

## 12 Az elektronikus iratok megőrzésének tétje

Anglia meghódítása után Hódító Vilmos parancsot adott, hogy a királysága javait mérjék fel és írják össze, így született meg 1086-ban a Domesday Book, a 11. századi Anglia páratlan forrása, ami a brit National Archives legféltettebb kincsei közé tartozik. Ennek mintájára 1986-ban a BBC elindította az ún. Domesday-projektet,<sup>102</sup> amelynek célja az volt, hogy dokumentálják – immár az informatika segítségével – a 20. század végi Egyesült Királyság

---

<sup>99</sup> Mark A. Greene: The Power of Meaning: The Archival Mission in the Postmodern Age. In: *American Archivist*, vol. 65. 2002, 45. old. Greene szerint az irattári paradigma egyik legjelesebb képviselője Luciana Duranti, az INTERPARES-projekt spiritus rectora.

<sup>100</sup> Mark A. Greene i.m. 53. old.

<sup>101</sup> Az InterPARES egyes tanulmányai azért marginálisan megemlítik a fondot és a magasabb szintű levéltári leírás fontosságát is, lásd pl. Luciana Duranti, Randy Preston (ed.): *InterPARES2: Experimental, Interactive and Dynamic records* (Associazione Nazionale Archivistica Italiana, 2008) 682.old.

<sup>102</sup> Lásd: <http://www.bbc.co.uk/history/domesday/story>

mindennapi életét. Az országot körzetekre osztották és iskolák, valamint helyi közösségek segítségével 150 ezer oldalnyi szöveggel és 23 ezer amatőr fotóval dokumentálták a környezetet. A dokumentációt az akkor még nagyon új, 12 hüvelykes optikai lemezen publikálták. Tervük szerint 25 évenként újra elvégzik ezt a munkát, így szemléltetve a bekövetkezett változásokat. 1999-ben azonban kiderült, hogy a 13 évvel korábbi projekt számítógépes háttere (hardver, szoftver) időközben eltűnt, és így az egész dokumentáció hozzáférhetetlenné vált. (Vilmos iratai viszont több mint 913 év után is jó és olvasható állapotban voltak.) Ekkor brit és amerikai szakemberek részvételével elindították a CAMILEON-projektet,<sup>103</sup> aminek keretében több évi munkával sikerült életre kelteni (emulálni) az 1986-os anyagot. Azóta a Domesday-projekt a digitális elavulás szimbólumává vált.

2002 és 2003 fordulóján az Egyesült Államok elnöke washingtoni hivatalának számítógépes szakemberei hardver- és szoftvercserét végeztek, aminek eredménye kb. 22 millió e-mail eltűnése lett.<sup>104</sup>

A fenti két példa jól szemlélteti, hogy milyen sokkoló következményekkel jár az e-iratok tömeges megsemmisülése. Az 1.2. fejezetben az iratok adminisztratív/gyakorlati, bizonyító és történeti értékét, ezek fontosságát taglaltam. A száraz szakmai fogalmak azonban alapvető értékeket, kockázatokat és következményeket takarnak.

Ha fontos levéltári iratok tűnnek el, az a társadalom jogbiztosító érdekét, valamint a múlt megismerésének a lehetőségét csorbítja, ami három fontos következménnyel jár: 1) az állampolgárok nem tudják érvényesíteni a jogukat; 2) a közszereplők (a hatalom) ellenőrzése sérül, ami visszaélésekre ösztönöz; 3) a történeti kutatásokat korlátozza, aminek számtalan következménye van az egymást követő nemzedékekre: politikaitól kezdve a társadalompszichológiáig bezárólag. Ezért nevezik a levéltárakat (ide értve az irattárakat is), pontosabban a levéltári iratokhoz való hozzáférést a demokrácia zálogának. A tekintélyelvű politikai rendszereket csalhatatlanul fel lehet ismerni: ha a hatalom igyekszik jogszabályokkal elzárni a polgárokat a levéltári iratok megismerésétől – indok persze mindig bőven van –, akkor ott baj van a demokráciával.

Az iratok megmaradásának és a hozzáférésüknek azonban van egy sokkal általánosabb, az egész emberiségre kiható hatása is. Az emberi civilizáció fennmaradásának és fejlődésének záloga, hogy a szerzett tudás megmaradjon és minél szélesebb társadalmi

---

<sup>103</sup> Lásd: <http://www.ariadne.ac.uk/issue29/camileon>

<sup>104</sup> Gewirtz, D.: *Where Have All the Emails Gone?*, Palm Bay, Fla.: Zatz. (2007). Idézi: Christopher J. Prom: *Preserving email* (DPC Technology Watch Report 11-01, 2011) 37.old. Elérhető: <http://www.dpconline.org/advice/technology-watch-reports>.

és földrajzi körben elterjedjen, mert így az egymást követő emberi generációk az előzők által megszerzett tudásra építve még több, újabb tudást halmozhatnak fel, így fenntartva a fejlődés folyamatát. Az emberi civilizáció történetében három olyan nagy impulzust konstatalhatunk, ami a tudás megőrzése és/vagy terjesztése területén hatalmas előrelépés volt: az írás, a nyomtatás és az informatika létrejötte, valamint elterjedése. Az írás – ami valószínűleg a világtörténelem legjelentősebb emberi alkotása – legfontosabb funkciója a tudás rögzítése volt, a nyomtatás pedig a megőrzés mellett a terjesztésnél is kiemelkedő szerepet játszott. Az informatika több tekintetben is új dimenzióba helyezte a fejlődést. Egyrészt ezzel lehetővé vált az emberi gondolkodás gépesítése, ami pl. a fizikai munkát végző gépekkel kombinálva a technikai fejlődést növelte ugrásszerűen. (Gondoljunk pl. a szoftver vezérelte megmunkáló robotokra az iparban.) A témánk szempontjából azonban fontosabb, hogy a tudás rögzítése és terjesztése hihetetlen mértékben hatékonyabb és kiterjedtebb lett, és a folyamatnak még csak a kezdetén járunk, a végét, a lehetőségek többségét még csak nem is látjuk.

A 2.3. fejezetben már érintettem ezt a kérdést, amikor az emberi kommunikáció és az emlékezet digitalizálásáról írtam. A felhalmozott emberi tudás, információ legnagyobb tárházai a levéltárak, a könyvtárak, a tudományos adatbázisok és egyéb gyűjtemények. Az informatika pár évtizedes története alatt óriási lökést adott sok tudománynak, így pl. a csillagászatnak, az atomfizikának, a biológiának, az orvostudománynak, a geológiának, amelyek ma már nem tudnának létezni adatbázisok, modellező szoftverek, szakmai hálózatok stb. nélkül.<sup>105</sup>

A fejlődés folyamatossága azonban nem feltétlenül biztos, mert amint az előző fejezetekben láthattuk, a megőrzés az informatika gyenge pontja, és minél nagyobb mennyiségű a digitalizált információ (a tudás), annál nagyobb lehet a veszteség. Ha a tudás elveszik, akkor az újabb generációknak nincs mire építeniük, sőt mivel elődeik tudását sem ismerik meg, megindul egy leépülés, hanyatlás. Persze nincs olyan veszély, hogy mindez az egyik napról a másikra bekövetkezik, de minden egyes megsemmisült, hozzáférhetetlen dokumentummal, adatbázissal, képpel stb. sérül az a tudásalap, amelyre az emberiség a fejlődését alapozhatja.<sup>106</sup>

---

<sup>105</sup> Egyes számítások szerint a tudományos adategyűttesek mennyisége – köszönhetően az egyre olcsóbb szenzorok adatgyűjtésének – évente megduplázódik! <https://www.sciencenews.org/article/preserving-digital-data-future-escience>

<sup>106</sup> Természetesen a fejlődésnek nemcsak technikai és tudományos feltételei vannak, az nagyban függ politikai, gazdasági, társadalmi és kulturális, sőt természeti feltételektől is, de ez kívül esik a vizsgálódási körünkön.

A tudományos világ ezt a veszélyt felismerte és aktívan fellépett. Az OAIS-t, a manapság már általánosan elfogadott e-levéltári szabványt, úrkutatási adatrendszerek archiválására fejlesztették ki, és a közreműködők között nincs egy levéltár sem. A Dublin Core főként könyvtárosok munkája, a PREMIS könyvtárosok, levéltárosok és egyetemi szakemberek összefogásából született, az amerikai Library of Congress sok szabványban fő szervező intézményként jelenik meg, a MOREQ-ben levéltárak, egyetemek, informatikai cégek vesznek részt és a sort folytathatnánk. A különböző kormányok, informatikai szakmai szervezetek, az Európai Unió az elmúlt évtizedekben rengeteg pénzt és erőfeszítést fordítottak az e-iratok megőrzésének elősegítésére. Tehát a levéltárak történetében egy páratlan összefogásnak lehetünk tanúi, és bár áttörés nem történt, azért ma már sokkal optimistábbak lehetünk, mint a kezdetekben. Az erőfeszítéseket azonban folytatni kell.