

EÖTVÖS LÓRÁND TUDOMÁNYEGYETEM

DOKTORI DISSZERTÁCIÓ

KÖRMENDY LAJOS

INFORMATIKA A LEVÉLTÁRBAN

(TÉZISEK)

2007

1 A kutatási feladat és a kutatás módszere

A fő kutatási feladatomban azt tekintetem, hogy felvázoljam az informatika hatását a levéltárelméletre és szerepét a levéltári gyakorlatban. A XXI. század elejére az informatika mindkettőt teljesen átformálta.

Az egyik legfontosabb kérdés számomra az volt, hogy létezik-e egy olyan szakmai állandó, ami egyaránt érvényes mind hagyományos, mind elektronikus környezetben, és ha igen, akkor ez miből áll. Ennek az állandónak, vagy másképpen szólva szakmai „magnak” a bemutatása több szempontból is lényeges. Egyrészt ez bizonyítja, hogy a levéltártannak vannak olyan állandó törvényszerűségei, melyek a tudományosság egyik fontos kritériumát jelentik. Konkrétan arról van szó, hogy a provenienciá- elv, amit a levéltárosok helyenként évszázadok óta alkalmaznak a gyakorlatban, és amit valamivel több mint száz évvel ezelőtt fogalmaztak meg először tudományosan, mennyire helytálló az informatika — különösen az elektronikus iratok — okozta kihívás után. Másrészt a „mag” feltárása segíthet a modern archivisztika legnagyobb dilemmájának megoldásában: a változékonyság talán az informatika legfontosabb jellemzője, ez azonban a levéltár ellen hat, ezért a levéltárosi közösség kitértően keresi a (technikai) állandót a változásokban. A disszertáció írásakor elsősorban arra törekedtem, hogy a szilárd, tartós pontokat emeljem ki.

Mivel az informatika a levéltárosi szakma minden szögletébe behatolt, ezért minden lényeges terület (levéltári leírás, digitalizálás, elektronikus iratok stb.) bemutatása szükségessé vált. Az általános körkép felvázolása mellett azonban kívánatos volt a lényegesebb pontok (pl. a levéltári információ) részletes elemzése is, hogy a „magot” megtaláljam, és ezért le kellett ásnia a levéltárban legmélyebb rétegeibe.

A levéltári informatika manapság már olyan sokrétű és olyan hatalmas tényanyagot jelent, hogy nincs szakmai polihisztor, aki minden kérdésben elsőrangú szakember lenne. A kutatásaimban — különösen azokon a területeken, ahol csak kevés személyes tapasztalatom van, például az elektronikus iratok tekintetében — főleg a nemzetközi szakirodalomra támaszkodtam, bár meg kell mondanom, hogy meglehetősen sok mű Budapestről elérhetetlen. Sokat segített viszont az internet, melynek segítségével — olykor rövidített formában vagy másodkézből — sok hiányzó cikket el tudtam olvasni. A szakirodalom mellett bőven merítettem a saját kutatásaimból és tapasztalataimból, melyek eredményeit az elmúlt közel húsz évben számos publikációban osztottam meg a szakmai közvéleménnyel.

2 A disszertáció felépítése

Bevezetőt követő második fejezetben azokat a társadalmi mozgatórugókat igyekszem felvázolni, melyek közvetlen hatással vannak a levéltárak ún. filozófiaváltozására. Mivel ezek az erők dialektikus kapcsolatban vannak az informatikával, annak működése és hatása a levéltárakra ezáltal jobban érthetővé válik.

A rövid harmadik fejezet a tervezésről és a levéltári informatikai stratégia kialakításának fontosságáról szól.

A negyedik és az ötödik fejezet témái szorosan összetartoznak. Az előbbi a levéltári információ elemével és a levéltári anyag szerkezetével foglalkozik. Az utóbbi az információ megjelenítési formájával, a levéltári leírást taglalja, és részletesen ismerteti a legfontosabb nemzetközi szabványokat, az ISADg-t, az ISAAR-t és az EAD-t.

A hatodik fejezet a levéltári informatika egyik legígéretesebb területével, a digitalizálással foglalkozik, elsősorban a képi digitalizálással, ami lehetővé teszi az iratok faksimile képének rögzítését

Az informatika talán az internetes alkalmazással befolyásolta legjobban a levéltárakat, ezt tárgyalja a hetedik fejezet: a levéltári intézmények hogyan tudják kihasználni a honlapok lehetőségeit, milyen követelményeknek kell eleget tenni a honlapok szerkesztésénél és működtetésénél.

A nyolcadik fejezet a levéltári informatika legakutabb problémáját az elektronikus iratok kérdését vizsgálja. Az elektronikus iratok számtalan problémát jelentenek mind az irattárakban, mind a levéltárakban, mert átalakul az irat fogalma, a forma, változik az elektronikus környezet, kérdéses az iratok hitelessége és megőrizhetősége. A fő kérdés az, hogy hogyan tudjuk úgy megőrizni az elektronikus iratokat, hogy azok hitelesek, integránsak és hozzáférhetőek legyenek évszázadok múlva is.

Az utolsó, kilencedik fejezet a levéltári rendszerekkel foglalkozik, először az általános követelményekkel, majd az Országos Levéltárban működő E-Archívum nevű rendszer elemzése révén mutatom be egy konkrét rendszer működését.

3 A disszertáció fontosabb megállapításai

3.1 Társadalom, levéltár, informatika

Az európai kultúrkörbe tartozó országokban a XX-XXI. század fordulója előtti pár évtizedben óriási tudati, társadalmi és gazdasági változások zajlottak le, melyek lényegesen átalakították a levéltárak filozófiáját (elvárások, értékek, gondolkodás). A filozófiai változásokat néhány kulcsszóval tudjuk jellemezni:

- *Popularizálódás.* A levéltárak az akadémiai vagy egyetemi tudósok számára rezervált kutatóhelyekből nagyszámú kutatót kiszolgáló, másolatokat nagy tömegben készítő intézményekké váltak. Az új kutatók nagyrészt laikusok, többségük a saját családjának történetét kívánja megismerni.
- *Átláthatóság és nyitottság.* Az iratok zárt periódusa egyre rövidült, és egyre több irat vált hozzáférhetővé. Általános lett a 30 éves zárolási idő, felülvizsgálták az iratminősítési szabályokat, és radikálisan csökkentették a titkosított iratok körét. Ugyanakkor a személyes adatok védelme egyre szigorúbban lett szabályozva.

A levéltárak érdekévé vált, hogy megismertessék magukat: sorra jelentek meg népszerű levéltári kiadványok, attraktív internetes honlapok, népszerűsítő rendezvények (kiállítások, nyílt napok), rendszeresen publikált éves jelentések.

- *Piaci szemlélet.* A levéltárakban is fokozatosan teret nyert a piaci-gazdasági szemlélet: fontos szempont lett a költséghatékonyság és a rentabilitás, nem ritka a levéltári intézmények átvilágítása sem, mert a fenntartó tőlük is megköveteli az ésszerű gazdálkodást.
- *Médiakultúra.* A levéltárak megfelelő témaválasztással népszerű kiadványok, honlapok és rendezvények révén széleskörű érdeklődésre tarthatnak számot.
- *Gyorsuló idő és információ-központúság.* A levéltárak legfontosabb kommunikációs eszközévé mára az internet vált. Sok levéltárban retrospektív feldolgozási munkát végeznek, azaz meglévő kiadványok, segédletek adatait táplálják be új informatikai rendszerekbe. Az elektronikus iratok megjelenésével jelentősen lerövidül a levéltárba adás határideje. A hagyományos iratok esetében általános 15-20 év immár nem tartható, mert addigra az elektronikus iratok számítógépes környezete annyira megváltozik, hogy az iratok egyáltalán nem vagy csak alig olvashatók.

Minden levéltárnak hosszú távon kell terveznie az informatikai tevékenységét. Az informatikai stratégiának illeszkednie kell a levéltár más feladatait (állományvédelem, szervellenőrzés, segédletkészítés stb.) átfogó intézményi stratégiájához.

3.2 Levéltári információ, levéltári anyag

A levéltári információ a következő elemekből áll:

- *Azonosító.* A levéltári anyag pontos egyedi jelölését teszi lehetővé, ilyen a törzsszám, a jelzet és a cím. A törzsszámoknak és a jelzeteknek kiemelkedő szerepük van az informatikai rendszerek működésénél.
- *Tartalom és struktúra.* Az iratban rögzített információ képezi az irat tartalmát, az irat, illetve az irat-együttes belső szerkezete, struktúrája, pedig azt mutatja, hogy a belső irategységek (iratcsoportok) hogyan vannak rendezve, milyen kapcsolat van közöttük.
- *Kontextus.* Az iratképzőre vonatkozó adatokat, információkat, valamint az iratképzővel, illetve az irattal-irategyüttessel kapcsolatos jogi, politikai, társadalmi vonatkozásokat jelenti.

A levéltári információ egyedi jellegét a struktúra és a kontextus adja.

Az információk, adatok irattá struktúrálnak, az iratok 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 (felülről lefelé haladva) fondszintet, állagszintet, sorozatszintet stb.

Az irategyüttesek közötti kapcsolatok iránya vagy vertikális, vagy horizontális lehet. A fondszint alatti kapcsolatokat a levéltári iratok mikrostruktúrájának nevezzük. Fondszint fölött is vannak kapcsolatok (fondcsoport/levéltár, fondfőcsoport/szekció, levéltári intézmény), ezek alkotják a makrostruktúrát. A mikro- és a makrostruktúra együttesen adja a levéltári anyag rendszerét.

A levéltárakban alkalmazott kutatási módszerek alkalmazkodtak a fent leírt rendszerhez. A levéltári információ tartalmi részét általában vertikális kapcsolatok révén éri el a kutató: a keresett iratot lépésről-lépésre, egyre szűkebb körben lokalizálva közelíti meg, tehát az általánostól jut el a specifikusig, ezt nevezzük indirekt információ-megközelítésnek. (A könyvtári információt általában közvetlenül, direkt módon — egy katalógusadat révén — érjük el.)

3.3 A levéltári leírás

A levéltári leírás az iratokban lévő, valamint azokkal kapcsolatos információk áttekintő és kivonatos rögzítése, ami történhet hagyományos vagy elektronikus módon. Levéltári leírást évszázadok óta készítenek levéltárosok, de az informatika ezen a területen is gyökeres változást idézett elő: integrációs erejének köszönhetően egyesíthetőkké váltak a különböző segédletek, elmosódtak a különbségek adatbázisok, valamint kéziratos és nyomtatott segédletek között, sőt, átjárás nyílt az integrált segédletek és a nyilvántartások között. A nemzetközi levéltárosi közösség az elmúlt tizenöt évben három olyan, világszerte elterjedt szabványt dolgozott ki, amely a levéltári segédletekben az adatok és információk egységes leírását teszi lehetővé bármely szinten (a fondtól az iratig): az ISADg, az ISAAR és az EAD.

Az ISADg koncepciója a proveniencia-elvnek megfelelően iratszinteken alapul, a leírt iratanyagnál pontosan bemutatja a vertikális és a horizontális kapcsolatokat, feltérképezi a levéltári anyag mikrostruktúráját (a fondtól az iratig), viszont nem említi a makrostruktúrát, pedig ott is sok, rögzítésre méltó információ van.

A mikrostruktúrát illetően véleményem szerint állag alatt nem érdemes szabványosítani a leírást. Egyrészt a levéltári anyag adatai túlságosan különbözőek, egy úrbéri összeírásban, közgyűlési jegyzőkönyvben, térképgyűjteményben vagy bírósági iratokban teljesen eltérő típusú adatok vannak, melyek más fajta feldolgozást (adatmezőket) igényelnek. Másrészt az adatok mennyisége az alacsonyabb szintek felé haladva mértani léptékben többszöröződik, és néhány speciális irategyüttest leszámítva főlegesen átfogó leírást adni ügyirat- vagy iratszinten, tökéletesen elegendő a karakterisztikus, releváns adatok listája vagy ezek kombinációja.

Az ISADg mezői pontosan lefedik a levéltári információ elemeit (azonosító, kontextus, tartalom és struktúra), és egyensúly van közöttük. A levéltári leírások (segédletek) történetében a három elem súlya koronként eltérő volt. A kezdetektől a XIX. század végéig a hangsúly a tartalom volt, régi leltárak, lajstromok többé-kevésbé csak az iratok tartalmát akarták prezentálni, a kontextus nem is létezett ezekben a segédletekben. Az azonosító elem szerepe a kezdetektől lényegében állandó volt, mert az azonosításhoz mindig szükség volt rá. A proveniencia-elv huszadik századi terjedésével a kontextus jelentősége fokozatosan növekedett. Az információs korban, különösen az elektronikus iratok miatt az azonosító elemet és a kontextust meg kellett erősíteni: nincs boríték, bélyeg, pecsét, hiteles aláírás, mindezeket helyettesíteni kell, és a helyettesítők többsége az azonosító kategóriájába tartozik. A kontextus-elem jelentősége is drámaian megnövekedett. A következőkben ismertetendő ún. sorozat-elv nagyon erős kontextus-reprezentációt követel meg, mert az iratok ki lettek szakítva az eredeti struktúrájukból, így a kapcsolatuk más iratokkal csak akkor rekonstruálható, ha az iratképzőről részletes leírás áll rendelkezésre. A kontextus azonban abszolút dominanciáját nem itt, hanem a szintén alább ismertetendő ún. poszt-kusztodiális elméletben érte el, melynek képviselői szerint a levéltári feldolgozó munka tulajdonképpen kontextus-információk szerkesztéséből áll.

Az ISAAR-ral nem írható le a levéltári anyag, nem szól a levéltári információ elemeiről, azonosítóról, tartalomról és struktúráról, mindezeket kimondottan az ISADg-re hagyja, az ISAAR csak az iratképző és az iratlétrehozás körülményeinek nagyon részletes leírására szolgál, azaz csak a kontextussal foglalkozik. Igen ám, de az ISADg-nek van egy külön kontextus-adatterülete, az ISAAR mégsem ennek a helyébe kíván lépni, hanem a két szabvány párhuzamos használatát és összekötését javasolja. Az ok az ausztrál levéltárosok által kifejlesztett ún. sorozat-elvben rejlik.

A proveniencia-elvben a levéltári anyag szervezése és rendszere a fondokon alapul, egy-egy fond pedig egy-egy iratképzőre (szervre, személyre) épül. A szervek azonban átalakulnak, megszűnnek, összeolvadnak vagy szétválhatnak, funkciójuk, kompetenciájuk változik, és a kurrens iratok szükség esetén az utódszervezethez kerülnek. A levéltáros emiatt kénytelen állandóan lezárni fondokat, illetve újakat

nyitni, lényegében azonos iratképző kör mellett. Az ausztrál levéltártudományban ezt kiküszöbölendő dolgozták ki a sorozat-elvet. Ennek fókuszában a sorozatnak nevezett iratfolyam áll, mely azonos eredetű iratokat jelent, amelyek együvé tartozását közös irattári rendszerük indokolja, vagy amelyeket együtt őriznek, mert azonos tevékenység eredményeként születtek, vagy mert azonos formátumúak és egy meghatározott funkcióra vonatkoznak. Az iratképző nem játszik meghatározó szerepet a sorozat-elv szerint kialakított irategyüttesekben. Például egy adott funkcióra (erdészeti gazdálkodás) vonatkozó iratokat az egymást váltó szervek fondjaiból külön sorozatba elkülönítik. Mivel egy sorozat több szerv teljes vagy részleges tevékenységének eredménye lehet, az iratképzőket és az iratokat — mivel aszinkronban vannak — nem célszerű a leírásnál szorosan együtt kezelni, azokat el kell választani. Egy sorozatot az ISADg-vel nem lehet leírni, pontosabban lehetséges, de csak rengeteg ismétléssel, mert ha egy fondképző iratait 6, 8 vagy 10 sorozatba osztottak szét, akkor a szerv leírását állandóan ismételni kellene minden egyes vonatkozó sorozatnál, ez pedig redundancia, ami az informatika alapelveivel ellentétes. Ezért volt szükség egy külön szabvány, az ISAAR kidolgozására, ezért nem volt járható út az ISADg kontextusának kibővítése.

Az EAD az internetre kifejlesztett, hardver- és szoftverfüggetlen SGML/XML leíró nyelvre épül, amelynek DTD-nek nevezett részében lehet definiálni azokat az elhatároló címkéket (tag-eket), amelyeket az adatfájlokban alkalmazunk. Az EAD tulajdonképpen egy levéltári DTD.

Az EAD más oldalról közelíti meg a levéltári leírás kérdését, mint az ISADg vagy az ISAAR. A leírás termékét elektronikus iratnak tekinti, és választ kíván adni az elektronikus iratokkal kapcsolatos kérdésekre. Az EAD-vel készült leírások szoftverfüggetlenek, időtállóak, könnyen konvertálhatók, szabványos adatelemekkel bírnak, modularizálhatók, és többféle felhasználást tesznek lehetővé. Ezzel a szabvánnyal lehetségessé válik, hogy minden létező iratanyagot leírjunk, a meglévő segédleteinket pedig átírjuk, azaz összes leírásunkat egy nagy keretbe foglaljuk, egyúttal a levéltár létrehozza a saját biztonságos elektronikus levéltárát, ugyanis a leírás eredménye is elektronikus iratnak számít.

3.4 Digitalizálás

Digitalizáláskor, illetve az elkészült digitális másolatokkal kapcsolatban a legfontosabb levéltári alapelvek és követelmények a következők: biztonság, rendezettség és hozzáférhetőség, metaadatok, hitelesség és rendszer.

A biztonsági követelményt, mely szerint a hordozó legyen tartós, és a rajta lévő információ hosszú időn keresztül olvasható, nehéz teljesíteni, mert egyrészt a hordozók eléggé instabilak, élettartamuk hossza bizonytalan, másrészt mert nagyon sérülékenyek, és a rajtuk lévő adatok „eltűnése” szabad szemmel csak ritkán észrevehető, harmadrészt mert a számítástechnikában nincsenek tartós szabványok, és ami ma olvasható, az nem biztos, hogy tíz év múlva is az lesz.

Biztonsági szempontból jó megoldás, ha a digitális képet egy bizonyítottan biztonságos hordozóról, mikrofilmről készítjük. A digitális másolatok megőrzésénél mindenképpen alkalmazni kell kockázatcsökkentő eljárásokat, ilyen például a megfelelő minőségű hordozó kiválasztása, az időnkénti átmásolás, a konvertálás.

Digitális képek esetében sokkal nehezebb az iratok közötti tájékozódás, mint az eredeti iratokban, ezért a digitalizálandó irategyüttesnek véglegesen rendezettnek kell lenniük. A rendezett iratokról először egy elektronikus segédletet kell készíteni, majd az elkészített képeket hozzá kell rendelni a segédlet megfelelő adataihoz, így a képek a segédlet segítségével lesznek kereshetők, listázhatók, böngészhetők.

Az elektronikus segédlet adatai tulajdonképpen metaadatoknak tekinthetők. (Metaadatok: adatok az adatról, esetünkben a digitalizált iratanyaggal és magával a fájjal kapcsolatos információk.) Különösen fontosak kép- és hangfájlok esetén, mert a szövegfájlokkal ellentétben ezeknél nem lehet szavakra, kifejezésekre keresni, tehát a nyilvántartásnak, keresésnek egyetlen eszköze a metaadat. Háromféle metaadat van:

- Leíró: az iratok azonosításához és az adatfeltáráshoz szükséges adatok.
- Adminisztratív: a digitális anyag használatához szükséges adatok (például technikai specifikációk vagy a készítés körülményeire, illetve kutatási korlátozásokra vonatkozó adatok).
- Strukturális: a digitális objektumok komponensei közötti kapcsolódásokra vonatkozó adatok (levéltári vagy technikai).

A legjobb az, ha minden egyes kép- vagy hangdokumentumot külön fájlba teszünk, és minden fájlhoz külön metaadat-rekord tartozik. Természetesen ez utóbbiakat össze kell kötni a vonatkozó fájlokkal, mert csak együttesen van értelmük. A metaadatok egy részét integrálhatjuk a fájlnevekbe. Képfájlok esetén erősen ajánlott, hogy a címük „beszédes” legyen, azaz tartalmazzák a digitalizált irat pontos jelzetét.

A legnagyobb gond a digitális másolatokkal, hogy könnyen manipulálhatók, és a beavatkozásnak közvetlenül nem marad nyoma. A jogi megoldás valószínűleg követni fogja a mikrofilmes modellt, azaz közvetett módon, a létrehozási folyamat szabályozásával, és megfelelő — főként automatikus — dokumentálásával válnak hitelessé a digitális másolatok. (Erről bővebben az elektronikus iratoknál lesz szó.)

A rendszerkövetelmény azt jelenti, hogy a létrejövő digitális információkat a levéltár informatikai rendszerének integráns részeivé kell tenni.

A levéltárak digitalizálhatják a segédleteiket, illetve az őrzött irataikat. A segédletek digitalizálásnak nagy előnye, hogy mivel minden kutatás segédletek tanulmányozásával indul, és folyik párhuzamosan, így oda koncentrálnak az összkutatások legnagyobb hányada, ezáltal viszonylag kevés számú oldal digitalizálásával lehet elérni magas használati arányt. Elsősorban olyan segédleteket érdemes képként digitalizálni, melyek szövegfájlként való konvertálása (OCR-rel) nem megoldható.

Bár manapság már százezer-, sőt milliószámra készülnek digitális másolatok, illúzió azt hinni, hogy a levéltárak teljes iratanyaga belátható időn belül digitális hordozóra fog kerülni. Mivel egy elektronikus segédlet készítése különösen munkaigényes szakmai feladat, érdemes az erőket egy-egy erősen kutatott, nagy irategyüttes digitalizálására koncentrálni.

A digitalizálást gondosan meg kell tervezni, ki kell alakítani az ezzel kapcsolatos stratégiát. Először meg kell határozni a szakmai célokat (például gyakran kutatott iratok kényelmesebb használatát biztosítani), utána fel kell mérni, hogy milyen eszközökkel lehet a célokat elérni (beruházás, személyi és pénzügyi feltételek), milyen módon történik az iratkiválasztás (például a kutatói rendeléseket automatikusan beillesztjük a digitális állományba), illetve el kell dönteni, hogy ki fogja végezni a szkennelést (például külső vállalkozó).

Előre meg kell határozni az általános szakmai és technikai követelményeket, például

- az iratkezelési követelményeket (lényegében ugyanazok, mint a mikrofilmezésénél),
- az alkalmazandó fájlformátumokat (őrzési célokra ajánlott a TIFF, kommunikációs célokra a JPEG),
- egyéb technikai követelményeket (élesség, kontraszt),
- a fájlnevrendszert (a képfájlok struktúrája, a „beszédes” fájlnevéadás elve),
- a kötelezően fölveendő metaadatokat,
- az alkalmazandó adathordozókat (őrzésre szalagos egység vagy DVD, napi használatra DVD vagy merev lemez).

3.5 Az internet alkalmazása

Az internet tulajdonképpen hierarchia nélküli hálózatok világhálózata, amely sok millió számítógépet kapcsol össze, melyek egyaránt lehetnek adatszolgáltatók vagy adatfelhasználók. Az internet állandó kétirányú kapcsolatot tesz lehetővé a világgal, azaz nem csak a levéltár, hanem a felhasználó is kezdeményezhet, közölhet adatokat, indíthat „tranzakciót”. Az egyszerű adatszolgáltatáson túl a levéltár megnyithatja virtuális kapuját az érdeklődő vagy a kutató előtt, aki az elektronizált adatokat, digitális képeket úgy nézheti, olvashatja, használhatja, mintha a levéltár kiállító- vagy kutatótermében lenne.

A levéltár internetes „ablakát”, a honlapot gondosan kell megtervezni, és nem kevés erőfeszítéssel kell működtetni. Ahhoz, hogy az internet (a honlap) megfelelően szolgálja a levéltárat, az alábbi követelményeknek kell megfelelnie.

A levéltár az internet által egy kommunikációs üzenetet küld a világnak, amelyben önmagáról ad képet, és meghatározza az elérendő célt. Például: olyan nyitott szolgáltató közintézmény, ahol komoly tudományos munka folyik, amely szívesen látja az érdeklődőket, mind a hazai, mind a külföldi kutatókat. A kommunikációs üzenetnek megfelelően kell alakítani a honlap formáját, struktúráját és tartalmát.

A honlapon lehet figyelemfelkeltő effektusokat és élénk színeket alkalmazni, azonban nem szabad túlzásba esni, mert az nem egyeztethető össze egy tudományos intézmény imázsával. Ugyanakkor a sok díszes irat-illusztráció azt a historizáló képet sugallja, hogy a levéltár elsősorban „rég és szép iratok” gyűjtőhelye, ahol a világtól elzártan komoly tudományos kutatómunka folyik — ez viszont nem egyeztethető össze a modern levéltárak képével.

A honlap szövegeinek szakszerűnek és közérthetőnek kell lenniük, a túlságosan tudományos szöveget a laikusok nem is értik. Itt sem szabad túlzásba esni, azaz stílusban sem szabad leszállni olyan alacsony intellektusú rétegek színvonalára, akik valójában nem képesek érdemi levéltári kutatásra.

A levéltárnak különböző felhasználói csoportokat (levéltárosokat, családkutatókat, nagyközönséget) egyszerre kell szolgálnia, ennek megfelelően különböző stílusokat kell alkalmaznia. A különböző csoportoknak szóló anyagokat már a menürendszerben célszerű elkülöníteni.

A menürendszer a honlap struktúráját mutatja. A teljes tartalmat érdemes erősen de nem túlságosan tagolni, mert a bonyolult menürendszerben könnyen eltéved a felhasználó. Arra kell törekedni, hogy a felhasználó mindig lássa, hogy a menürendszerben éppen hol van.

A levéltári honlapon a következő blokkokat érdemes kialakítani:

- A levéltár bemutatása (levélcím, telefon, e-mail, fax, hol van a levéltár, hogyan közelíthető meg, nyitvatartásra vonatkozó adatok, mi a levéltár feladata).
- Kutatással kapcsolatos információk (a kutatóterem helye, nyitvatartási ideje, a kutatás feltételei [kutatási szabályzat], másolatkészítési szabályok, árak).
- Információk kezdő és laikus kutatóknak (főleg család- és helytörténeti kutatóknak módszertan, példák).
- Levéltári iratok leírása (elektronikus segédletek, on-line adatbázisok).
- Dokumentumgyűjtemények, kiállítások (a levéltár anyagából a szélesebb publikumnak szóló, népszerűsítő tudományos tanulmányokkal ellátott, illusztrált dokumentum-gyűjtemény, illetve kiállítási anyag).
- Publikációk (tanulmánykötetek, periodikák, segédletek).
- Szakmai információk (a szakmai közösség vagy a belső dolgozók számára szabályzatok, módszertani művek, útmutatók és egyéb segédanyagok).

3.6 Elektronikus iratok

Az elmúlt 100-120 évben a levéltárak szakma fejlődésében 50-60 évenként történtek olyan minőségi változások, melyek a levéltárban lényegi részeit érintették. A XIX. század végén az európai polgári társadalmak fejlődése megkövetelte, hogy a köz- és magániratok megőrződjenek, és visszakereshetők legyenek. A szakmai közösség válasza a proveniencia-elv megfogalmazása volt, ami lehetővé tette, hogy a levéltárak átfogóan begyűjtsék és feldolgozzák a szervek, szervezetek és jelentős magánszemélyek iratait. A következő nagy változás a XX. század közepére datálódik, amikor elsősorban az állam egyre kiterjedtebb szerepvállalása miatt robbanásszerűen megnőtt a keletkezett és feldolgozandó iratok mennyisége. Ekkor új levéltári módszerek sorozatát konstatálhatjuk az iratértékelésben, a selejtezésben, megjelent a mikrofilmezés és az állományvédelem. A harmadik választóvonalhoz most érkezünk el az elektronikus iratokkal.

A hagyományos és az elektronikus irat között sok különbség van:

- Az elektronikus irat csak bonyolult segédeszközökkel (hardver, szoftver) olvasható.
- Az elektronikus irat sokkal szélesebb spektrumú, mint a hagyományos (szövegszerkesztett iratok, táblázatok, adatbázisok, kép- és hangdokumentumok).
- Az elektronikus iratban az információ elválik az adathordozótól, az adatok szétszórta tárolódnak, csak a megfelelő szoftverrel ellátott számítógép tudja, hogy mi hova tartozik.
- Az elektronikus iratban a forma is gyakran elválik a tartalomtól. A formát befolyásolja a hardver és a szoftver egyaránt.
- Az elektronikus iratban a tartalom is fragmentált lehet. Egy irat fizikailag 5-10 különböző fájlban lehet tárolva, és csak a szoftver „hozza össze őket” a képernyőn.

- Az elektronikus irat jellemzően sokkal bonyolultabb struktúrájú, mint a hagyományos (lásd az internetes linkek szerepét).
- Hagományos fondokban az iratok általában követik a szerv szervezeti felépítését, az elektronikus iratoknál a szervezeti egységekhez köthető határok elmosódottabbak: a különböző osztályok munkatársai közös adatbázisokból dolgoznak.
- Az elektronikus iratoknál gyakran elmosódnak a fondhatárok: a szervek közös adatbázisokat vagy egyéb iratforrásokat használnak.
- Az elektronikus iratoknál — a használat hardver- és szoftverfüggősége, az iratok fragmentáltsága és bonyolultsága, a nehéz tájékozódás miatt — megjelentek és kiemelt szerepet kaptak a metaadatok.
- Az elektronikus iratok intellektuálisan nagyon instabilak, gyorsan, állandóan változnak, a hagyományosak sokkal hosszabb életűek.
- Az elektronikus iratok fizikailag is instabilabbak és sérülékenyebbek, mint a hagyományosak: a lemezek és szalagok pár év alatt lemágneseződnek, az optikai adathozdozók egy karcolással tönkretelhetők.

Az elektronikus iratokban az információ-elemek súlypontja eltérő a hagyományos iratokhoz képest. A kontextus-elem súlya megnövekedett, a fogalma is kibővült, mert nem csak a szervezetre vagy személyre vonatkozó jogi, közigazgatási, gazdasági, történelmi stb. kiegészítő információkat kell érteni alatta, hanem informatikai összefüggéseket is.

Hasonlóképpen megnőtt a struktúrára vonatkozó elem súlya, mert pontosan rögzíteni kell a kapcsolódó iratok (rekordok) és irategyüttesek (táblák, fájlok) közötti összefüggéseket. A kontextus- és a struktúraelem két eszközzel erősíthető meg: részletes dokumentációval és metaadatokkal.

Az elektronikus iratok archiválásának főbb problémái a következők:

- Hardverfüggőség (a gyors változások és az eltérő szabványok miatt a régi elektronikus iratok az új berendezésekkel már nem olvashatók).
- Szoftverfüggőség (a szoftverek választéka rendkívül nagy, túlnyomó többségük egymással nem kompatibilis, a hardvernél is gyorsabb szabványváltozás miatt rendkívül gyors az elavulás).
- Zárt kódok, jogvédelem (a fájlformátum-kódok többnyire nem nyilvánosak, így a kívülálló szoftverszakemberek hátrányban vannak a működtető programok írásánál; a szoftverek többsége jogvédett, a szervek nem adhatják át azokat a levéltárnak az elektronikus iratokkal együtt).
- Hitelesség (a forma és a tartalom elválása, a forma változása, a tartalom könnyű, és nyom nélküli változtathatósága, az elavulások miatti konvertálások miatt az elektronikus iratok hitelessége gyengül. Hogy egy irat hiteles maradjon, meg kell őrizni az azonosságát [egyediségét], integritását és használhatóságát. Ezeknek a követelményeknek egy sor szabály betartásával és biztosíték kiépítésével lehet megfelelni, mint például az iratképzés és használat szabályozottságával, megfelelő dokumentálással [naplózás], jogosultsági rendszerrel, metaadatokkal, rendszeres ellenőrzéssel és migrálással).

A fenti problémákra nem jelentenek megoldást az üzleti életben elterjedt irat- vagy tartalomkezelő rendszerek (records vagy content management systems). Ezek a különböző formátumú fájlokat (iratokat, irategyütteseket) egy rendszerbe foglalják és kezelik, és kb. egy évtizedre visszamenően tartalmazzák a gyakran használt programok olvasó- és böngésző funkcióját, az ennél hosszabb idő viszont meghaladja az üzleti szektor horizontját. Másrészt az irat- és tartalomkezelő rendszerek az egyszerű, egyedi iratokra és táblákra koncentrálnak, de adócsak maradnak az iratrendszerekben lévő összefüggések megjelenítésével.

A szerveknél, személyeknél keletkező elektronikus iratok szabályozásának igénye elsősorban a hitelesség és a hozzáférhetőség miatt vált szükségessé. A fejlett elektronikus közigazgatással bíró országokban kidolgozták azokat a követelményeket, modelleket vagy szabványokat, melyek az egységesítést szolgálják. Ezek általában az iratkezelő rendszerekre vonatkoznak, mint például az INTERPARES vagy MOREQ.

Az INTERPARES az elektronikus iratból indul ki, ezt vizsgálja, szedi apró alkotó elemekre, és mindent abból ítél meg, hogy mennyire szolgál bizonyítékként (elsősorban adminisztratív és jogi szempontból). Bár arra törekszik, hogy az irat teljes életciklusát lefedje — ellentétben a MOREQ-vel, amely a levéltári ciklussal nem foglalkozik —, az INTERPARES követelményrendszere nagyon statikus, nehezen áttekinthető, egysíkú a munkamódszere és a „hitelességi csőlátása” miatt. A MOREQ viszont a funkcioná-

litásból indul ki, ezért nem statikus, áttekinthető, rugalmas és sokoldalú. Az INTERPARES alkotói az ún. irattári paradigmát képviselik, ők az iratot csak bizonyítéknak tekintik, és emiatt az irat bizonyító erejét, kontextusát, valamint a formai-strukturális jegyeit abszolútizálják.

A metaadatok „világszabványává” az ún. Dublin Core vált, a levéltárakban azonban csak módosításokkal, kiegészítésekkel alkalmazható. Levéltári szempontból ugyanis komoly hiányosságai vannak, mert egyrészt az alkalmazott terminus technikusok egy része csak módjával felel meg a levéltári-irattári fogalmaknak, másrészt nem veszi figyelembe az irattári-levéltári iratok legfontosabb jellemzőit: a hierarchiát, a struktúrát, a szinteket, harmadrészt nem követi az iratok életciklusát, főleg nem különíti el az iratok irattári és a levéltári fázisát.

Az elektronikus iratok fájlformátumainak szabványosítása kulcskérdés a levéltárak számára. A formátumok éppen olyan gyorsan változnak, mint a szoftverek, egyébként is hardver- és szoftverfüggő a megjelenítésük és használatuk. Nincs egységes álláspont a nemzetközi szakirodalomban és gyakorlatban az ajánlott fájlformátumokat illetően, legfeljebb általános elméleti szinten:

- jelenítsenek meg minden fontosnak tartott információt és kapcsolatot;
- legyenek nyílt kódúak, és nemzetközi vagy nemzeti szabvány által elismertek;
- legyenek hosszú életűnek és széles körben elterjedtek;
- legyenek hozzáférhetőek és további konvertálásra alkalmasak;
- legyenek hardver- és szoftverfüggetlenek;
- legyenek alkalmasak az eredeti rendszerbe való visszakonvertálásra.

Jelen ismereteink szerint a fenti követelményeknek többnyire (néha erős kompromisszum árán) az alábbi formátumok felelnek meg:

egyszerű szövegek: RTF, DOC, PDF, TXT

strukturált szövegek: SGML, XML

táblázatok: XLS, CSV, SGML, XML

adatbázisok: SGML, XML

képek: TIFF, JPEG

mozgóképek- és hangdokumentumok: MPEG-1 és MPEG-2, Quick Time, Windows Media Video, illetve MP3, Apple Audio Interchange Format.

A hosszú távú levéltári megőrzés tekintetében négy olyan stratégia van, amit a nemzetközi szakirodalom jelenleg alternatívának (vagy párhuzamosan alkalmazandónak) ismer el: a migrálás, az emulálás, a technológiai megőrzés és a poszt-kustodiális levéltár.

Migrálás: a fájlokat a régi, elavult formátumokról új szabványoknak megfelelő formátumokra konvertáljuk. Ez a legelterjedtebb megoldás a világ élenjáró levéltárai között. Hátránya, hogy elkerülhetetlen adatvesztéssel jár. Nem mindegy azonban, hogy milyen mennyiségű és minőségű adatokat veszítünk. A tartalmat, illetve a lényeges funkciókat reprezentáló kapcsolatokat nem vagy csak minimális mértékben érheti veszteség, a formai megjelenítéssel kapcsolatosan már lehetünk nagyvonalúbbak. A sokszoros migráció azonban halmozott adatvesztést eredményezhet, ezért érdemes azt a lehető legritkábban, de az adatbiztonságot még nem veszélyeztető intervallumokként elvégezni.

Az elterjedt, szabványos, nem jogvédett, platformfüggetlen, az alkalmazási szoftverektől jól függetleníthető fájlformátum stratégiai előnyt biztosít. Jelen tudásunk szerint egyetlen ilyen formátum létezik, az SGML/XML.

Emulálás: az eredeti hardver- és szoftverkönyezetet (operációs rendszert) egy külön, erre a célra készült szoftverrel imitáljuk, így az adatokat az eredeti alkalmazói programmal úgy futtathatjuk, mintha az eredeti környezetben tennénk.

Az emulálás rendkívüli informatikai felkészültséget igényel, hiszen az eredeti elektronikus iratok, valamint az alkalmazói szoftver mellett meg kell őrizni az eredeti operációs rendszert és azokat a fájlokat, melyek az alkalmazói szoftver működéséhez szükségesek. Emellett le kell írni az ún. emulátor-specifikációkat, azaz az eredeti hardver-platform minden olyan attribútumát (például sebesség, display-attribútumok, eszközök és perifériák, multi-user aspektusok), ami a jövőben is lehetővé teszi az aktuális emulátor (illesztő szoftver) programozását. Ezenkívül rögzíteni kell a metaadatokat és egyéb annotációkat.

Végül az emulációs stratégiát alkalmazó levéltáraknál érvényesek mindazok a hátrányok, amik az eredeti szoftver átvételével kapcsolatban fennállnak:

- A sok alkalmazói szoftver használatának megtanulása a mennyiség növekedésével hatványozott nehézséget okoz mind a levéltárosoknak, mind a kutatóknak.
- A hardver és a szoftverkörnyezet (operációs rendszer) változásával az emulálást újra meg újra el kell végezni.
- A jogvédett alkalmazói szoftverek gyakran licencdíjasak, ami halmozódás esetén rendkívül sok pénz rendszeres kiadását eredményezi.

Technológiai megőrzés: a levéltár nem csak az eredeti fájlokat és a kezelő szoftvereket veszi át és őrzi meg, hanem a hardvert is. Ez a stratégia napjainkban teljesen visszaszorult, ugyanis megoldhatatlan gondot jelent az állandóan növekvő géppark üzemeltetése és karbantartása. Mindezek ellenére bizonyos esetekben van létjogosultsága a technológiai megőrzésnek. Érdekes olyan elavult hardverelemeket megőrizni, melyek használata szükséges lehet későbbiekben felbukkanó elektronikus iratok olvasására.

Poszt-kustodiális levéltár: a levéltár nem őrzi elektronikus iratokat, mert azok a szervek őrzetében maradnak, a levéltár feladata csak az, hogy internet segítségével intellektuális kontrollt érvényesítsen. Eme elmélet ellen komoly szakmai és etikai kifogások fogalmazhatók meg. A levéltár ezzel beismeri, hogy képtelen megoldani az elektronikus iratok problémáját, és áthárítja a felelősséget, valamint a munkát az iratképző szervekre. Másrészt az iratokat komoly veszély fenyegetné, mert a szervek nem érdekeltek az archív, számukra már fölösleges elektronikus iratok karbantartásában és konvertálásában.

Az elektronikus iratok állományvédelme az egyik legfontosabb feladata a levéltáraknak. Ki kell választani azokat a hordozókat, amelyeken az elektronikus iratokat tárolni fogja. A kiválasztásnál a következő szempontokat kell mérlegelni: stabilitás, élettartam, elterjedtség (szabványosság), kapacitás, kezelhetőség, adatmódosíthatóság (a hitelesség miatt fontos). Általánosságban elmondható, hogy a mágneses szalagok (DAT/DDS, illetve DLT/SDLT) vagy 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 adatok módosíthatók, viszont széles körben használatosak, és rendkívül nagy a kapacitásuk. Az optikai lemezek stabilak és hosszabb élettartamúak (igaz, mechanikus hatásra nagyon sérülékenyek), általánosan elterjedtek, a kapacitásuk jelenleg még kicsi (de rohamléptekkel nő), könnyen kezelhetők, az adataik pedig nem módosíthatók (kivéve az újraírható CD-ket és DVD-ket).

Ajánlott több példányban őrzeni az elektronikus iratokat: a biztonsági példányt egy távoli helyen, a mesterpéldányt optimális raktári körülmények között tartjuk, a munka- vagy kommunikációs példányt pedig átlagos körülmények között kezelhetjük. A saját elektronikus iratoknál fontos az ún. back-up-stratégia, azaz annak a meghatározása, hogy a gyorsan változó tartalmat milyen rend szerint, milyen időközönként rögzítsük.

A mágneses adathordozók használatát minimalizálni kell — kivéve a merevlemezeket —, a magas hőmérséklet és relatív páratartalom illetve ezek ingadozása kerülendő, a fénytől, különösen napfénytől védve kell azokat tartani.

Magyarországon még nem vesznek át elektronikus iratokat a levéltárak, ezért ez idáig csak a szerveknél keletkező elektronikus iratok szabályozásában sikerült eredményt elérni. A legfontosabb jogszabályok a következők: a 7/2005 IHM rendelet a nem állami és önkormányzati szervek digitális archiválásáról, a 12/2005 IHM (az ún. formátum-) rendelet, ami a közigazgatási eljárásokban keletkezett iratok technikai szabályait részletezi, a 13/2005 IHM rendeletet, mely azt szabályozza, hogy mikor válik hitelessé a digitalizált másolat, valamint a 335/2005. sz. kormányrendelet, ami a közfeladatot ellátó szervek iratkezelését írja elő.

A fenti jogszabályok közös vonása, hogy az elektronikus iratokat nem rendszerben vizsgálják, illetve szabályozzák, ezért az iratok hitelességének biztosítását kénytelenek teljesen az elektronikus aláírássokra bízni. A formátum-rendelet nagy hiányossága, hogy hiányoznak belőle a strukturált szövegek, táblázatok, adatbázisok, mozgóképek, hangdokumentumok formátum-előírásai, ami gyakorlatilag szabad kezet ad a szerveknek a formátum-alkalmazásnál.

3.7 Levéltári rendszerek

Egy levéltári rendszert nagy gonddal kell tervezni, benne a lehető legtöbb adategyüttest és funkciót célszerű integrálni az egész kijelölt vertikumban, mindennek előre kijelölt helye legyen (kép, metaadat, irat stb.), így gyarapodásakor minden automatikusan a helyére kerül. Érdemes úgy tervezni, hogy újabb feladatok, funkciók (adminisztráció, nyilvántartás, szkennelt iratanyag stb.) könnyen beilleszthetők legyenek. Egy jól felépített rendszer legnagyobb előnye, hogy nincs adatismétlés, és minden információ naprakész. Egy központi adatbázis köré célszerű építeni a rendszert.

A 2002-ben megjelent 10/2002. NKÖM rendelet előírásának megfelelően az Országos Levéltár által kidogozott E-Archívum nevű rendszer csak részben tud megfelelni az informatikai és a levéltártani követelményeknek. Meglepő módon hiányzanak vagy nagyon kezdetlegesek egyes informatikai alapfunkciók, mint például a nyomtatás, az export vagy import. Emellett komoly levéltártani hiányosságokat is tapasztalhatunk.

Törzskönyvi címszó alatt a törzskönyv, valamint a fond- és állagjegyzék egy furcsa keverékét kapjuk. Benne állag alatti „törzslapokat” is találunk, ami teljesen szabálytalan mind a törzskönyv, mind a fond- és állagjegyzék esetében, ráadásul a rendszertől nem tudhatjuk meg, hogy mi volt az utoljára kiadott törzsszám. Az áttekintő törzskönyvi struktúrának nevezett adatok nem felelnek meg a fond- és állagjegyzék követelményének sem, mert egyrészt hiányoznak az évköri és terjedelmi adatok, másrészt az állag alatti rekordok ide sem illenek.

Az iratszintek (állag, fond, fondcsoport stb.) önkényesek és zavarosak: ismeretlen fogalmakkal operál (mint például állagközépcsoport), és teljesen összekuszálja a levéltári rendszert: például a fondfőcsoport/szekció és a fondcsoport/levéltár közé beteszi az ún. levéltárfőcsoport és a mikrofilmes levéltári intézmény szintjét.

Nagy problémát jelent, hogy az eredeti iratokkal együtt egy táblába tették a fényképgyűjteményt és a filmtárat. Mivel ezek a másolatok őrzik az eredeti iratok (amiről a másolat készült) szintadatait, így a hierarchia felborul, például fondfőcsoport alatt szerepelhet ország- vagy levéltári intézmény-szint. Hasonló gondot jelent az elkülönült adatállományok (például: úrbéri tabellák) beolvasztása, ráadásul az eltérő adatmezők nem illeszkednek egy általános nyilvántartás adattartományába, ezért azokat a programozó kénytelen volt az általános leírás című mezőbe tenni.

Ezekben az általános leírásokban azonban az E-Archívum nem tud keresni, így több ezer oldalnyi információ gyakorlatilag rejtett marad a felhasználó számára. A címmezőben lehet keresni, de a találatok kiírása jelzetek sorrendjében történik, ami nagy hiba, az eredmény zűrzavaros lesz, a tájékozódáshoz legalább látni kellene a hierarchiát.

Listázási funkció nincs a rendszerben, pedig levéltári adatbázisoknál ez alapfunkciónak számít.

Az iratanyag tartalmának leírása az E-Archívumban Fondismertetés címen szerepel (bár nem csak fondoknál alkalmazható). Két részből áll, az első elkülönült mezőket kínál fel leírásra, a másik csak egy általános mezőt kínál fel. Az első némileg hasonlít az ISADg mezőire. Az elkülönült mezők struktúrát alkotnak, és szabványosítást tesznek lehetővé, de csak akkor, ha azt a rendszer megfelelő kereső-szűrő- listázó-, export- és import-funkciókkal támogatja, ugyanis csak így lehet kiaknázni a lehetőségeket. Mivel az E-Archívumban ez nincs így, ezért a fondleírások holt zárványként szerepelnek a rendszerben.

Az E-Archívum fontos modulja a szervnyilvántartó. A NKÖM-rendelet előírja, hogy a szervek jogelőd-jogelődjeit is nyilván kell tartani. Az E-Archívum túlmege ezen, és a jogutódokra is tartalmaz rovatot. Ez a plusz szolgáltatás mind jogilag, mind szakmailag erősen aggályos. A jogszabály csak a még létező szervek nyilvántartását írja elő, lévén, hogy a szervnyilvántartás az iratkezelés felügyeletének eszköze. Ha jogutódot is nyilvántartunk, akkor a megszűnt szervek is a nyilvántartás egyenrangú részét képezik, ami ellentmond a jogalkotó szándékának. Bizonyítja ezt, hogy a megszűnt szerveknél egy sor adat (a szerv címe, elérhetősége, a kapcsolattartó személy adatai stb.), értelmetlenné válik. Kétségtelen, hogy szükséges a megszűnt szervek nyilvántartása, de a két dolgot el kell választani, ami programozás szempontjából egyszerű feladat lenne. Jelenleg az élő és a megszűnt szervek nem különíthetők el, sőt azt is csak kikövetkeztetni tudjuk, hogy egy adott szerv melyik kategóriába tartozik (ha van jogutód, akkor megszűnt).