

Born digital workflow tervezése a Petőfi Irodalmi Múzeum Digitális Bölcsészeti Központjában

Kalcsó Gyula

Petőfi Irodalmi Múzeum Digitális Bölcsészeti Központ (Budapest)

Eszterházy Károly Egyetem (Eger)

kalcsogyula@dbk.pim.hu

The Petőfi Literary Museum's Digital Humanities Centre, established last year, is in the process of creating a digital humanities platform, dHUpla. One of its sub-areas is the planning of the management of born digital materials. The development of a procedure for born digital materials is an urgent need: in addition to the existing collection items in the PIM, it is expected that in the future more and more content of this kind will be created, which will need to be managed professionally. The presentation will describe the research work carried out and the details of the workflow that is being outlined. It will present the workflow developed on the basis of the OAIS model, which covers the process from ingestion of such material to making it searchable and publishable. It covers not only the more manageable text file formats of born digital content, but also the more difficult tasks of emails or social media sites, and how to professionally extract and archive data from external media.

Keywords: born digital, OAIS, BagIt, dHUpla, digital humanities platform

1. A born digital objektumok gyűjteményi kezelése Magyarországon aránylag új terület. Részletesen kidolgozott born digital workflow-val egyelőre csak a Magyar Nemzeti Levéltár büszkélkedhet, az ő gyűjteményi profiljuk azonban egy kissé speciális, a levéltári iratok egy jó része kötelezően szolgáltatandó az iratképzők által, ezért 2016-tól kormányrendelet szabályozza a levéltári elektronikus iratbeszolgáltatás rendjét¹. A kulturális közgyűjtemények born digital anyaga más jellegű, már ezen a területen is van kialakult gyakorlat, de elméletileg és tudományosan kevésbé megalapozott. A nagyobb közgyűjteményi szereplők természetesen rendelkeznek a born digital anyagokra vonatkozó, többé-kevésbé szabályozott eljárásrenddel: léteznek elektronikus könyvtárak, intézményi repozitóriumok, ahol kezelnek born digital objektumokat. Ugyanakkor ezek kialakításának az elvei, a használt módszerek, szoftverek közel sem egységesek, ebből fakadóan sajnos sok esetben egymással kölcsönösen nem (teljesen) átjárhatóak, valamint a köztük lévő szemantikus kapcsolatok sem gyakorlatiak.

A Petőfi Irodalmi Múzeum Digitális Bölcsészeti Központja 2020-ban alakult meg, alapító okirata szerint "a hazai közgyűjteményekben zajló digitalizáció, tartalom- és szolgáltatásfejlesztések támogatására, koordinálására és dinamizálására" jött létre².

- 1 Magyar Közlöny, „34/2016. (XI. 30.) EMMI rendelet az elektronikus formában tárolt iratok közlevéltári átvételének eljárásrendjéről és műszaki követelményeiről”, elérés 2021. május 15., <https://magyarkozlony.hu/dokumentumok/afc7a7bc8a8b8c9e22db85179ec5932bc5fa05e9/megtekintes>.
- 2 <https://pim.hu/hu/digitalis-bolcseszeti-kozpont/rolunk>, elérés 2021. május 15.



A PIM-DBK kettős indíttatásból szándékozik a born digital objektumkezeléssel intenzíven foglalkozni: szükséges egyfelől a PIM gyűjteményezési tevékenységét támogatnia, a már meglévő gyűjteményi elemek feldolgozását elvégeznie, másfelől egy elméletileg és tudományosan megalapozott kulturális közgyűjteményi jó gyakorlat kialakítását is célul tűzi ki.

Az SZMSZ-e szerint „a PIM gyűjtőköre kiterjed a magyar irodalom egészének tárgyi, képi (képzőművészeti, fotó, videó és film), valamint kéziratos és nyomtatott, digitálisan keletkezett, illetve könyv formában található emlékeire, ill. hangzó dokumentumaira, valamint a magyar színházművészet és színházi kultúra muzeális tárgyi, képi, írásos és egyéb (hang-, film-, digitális anyag, stb.) emlékeire és az ezzel összefüggő képzőművészeti alkotásokra; illetve a tagintézmények vonatkozásában a kulturális javak mindazon körére, amelyet a tagintézmények muzeális működési engedélye gyűjtőkörként meghatároz. [...]

Fő feladata a gyűjtőkörébe tartozó muzeológiai forrásanyag felkutatása, gyűjtése, raktári megőrzése, műtárgyak kölcsönzése, visszazorolása, szakszerű nyilvántartása, kezelése, állagmegóvása és védelme, restaurálása; továbbá tudományos feldolgozása és rendezése, mindezek kiállításokon és más formákban történő bemutatása, a közművelődését segítő hasznosítása.”³

A born digital objektumokkal kapcsolatos összetett feladatrendszer tehát egyfelől a közgyűjteményi hatáskörből adódik, másfelől viszont felszínre hoz olyan megoldandó problémákat és elvégzendő teendőket is, amelyek korábban nem voltak a fókuszban. Ennek a szerteágazó tevékenységnek a támogatására szükséges a PIM-DBK-nak kidolgoznia egy eljárásrendet, amely megfelel a Közgyűjteményi Digitalizálási Stratégiában⁴ megfogalmazott követelményeknek, összhangban van a PIM elvégzendő feladataival, és amelyet más közgyűjteményi szereplők számára is ajánlhat (szolgáltatásként is) abban a reményben, hogy létrejön egy egységesebb, átgondoltabb, szabványosabb, nemzetközileg is elfogadott normáknak megfelelő born digital-kezelési gyakorlat. A kidolgozandó eljárásrend a DBK által fejlesztett dHUpla⁵ nevű digitális bölcsészeti platform workflow-jának része lesz.

2. A born digital kifejezés digitálisan született, eleve digitálisan létrejött objektumokat⁶ jelöl, azaz tulajdonképpen minden olyan fájlt vagy több fájlból álló egységet, amely nem valamely korábban már analóg módon létező objektum digitalizálásával keletkezik. Ugyanakkor a born digital nem jelenti azt, hogy ezek az objektumok kizárólag digitálisan létezhetnek (készülhet analóg másolat róluk), viszont a kezelésükkel kapcsolatos problémák a digitális létmódjukból adódnak. A born digital objektum legalább egy, de sokszor több fájlból áll, amelyek lehetnek könyvtárakba szervezve. Fontos és problémás

3 https://pim.hu/sites/default/files/book/file/pim_szmsz_2020.10.14.pdf, elérés: 2021. május 15.

4 <https://digitalisjoletprogram.hu/hu/tartalom/kds-kozgyujtemenyi-digitalizalasi-strategia>, elérés: 2021. május 15.

5 A Digital Humanities Platform kifejezésből alkotott betűszó.

6 A digitális objektum kifejezés a Fehér könyv ajánlása a PREMIS alapján. Bánki Zsolt és Kómár Éva, szerk., *Fehér könyv. Módszertani útmutató a közgyűjteményi kulturális örökség digitalizálásához és közzétételéhez* (Emberi Erőforrások Minisztériuma, 2019), 123. https://ommik.hu/media/attachments/2019/12/09/fehr_knyv.pdf. Elérés: 2021. május 15.

esetet jelentenek azok a “zárványok”, amelyek elavult adathordozókon (esetleg elavult formátumban) találhatóak, ezek megmentése elemi közgyűjteményi érdek.

A born digital objektumok kezelésének egyik fő nehézsége a heterogenitásukban rejlik. Ez jelenti egyrészt azt, hogy nagyon sokféle digitális objektumtípus lehetséges, másrészt egy-egy digitális objektum is állhat informatikailag eltérő típusú összetevőkből. Ugyanakkor kezelésük nehézségei közé tartoznak a digitális archiválás általános problémái, amelyek nem csak a born digital objektumokra vonatkoznak: a nagy mennyiség kezelése, a változatosságból adódó problémák megoldása, az elavulás kiküszöbölése, az értelmezhetőség fenntartása.

2.1. A born digital tárgykörébe tartozó objektumok meghatározása nemzetközi szinten sem egységes. A Közgyűjteményi Digitalizálási Stratégia⁷ alapidokumentumának számító *Fehér könyv*⁸ szerint “a digitális objektumok lehetnek szövegek, álló- és mozgóképek, hanganyagok, adatbázisok, adathalmazok, 3D objektumok és a felsoroltak konténerai”.⁹ Ez alapvetően tartalmi kategorizáció, a közgyűjteményi válogatás szempontjainak meg is felelhet, de a digitális megőrzésben általában inkább a fájlformátum alapján szokás osztályozni a digitális objektumokat. Ugyanakkor a fájlformátumok szerinti kategorizáció sem visz közelebb a born digital meghatározásához, hiszen bármilyen típusú fájl(ok együttese) lehet born digital tartalom. A born digital objektumok azért jelentenek speciális problémát, mert vannak olyan tulajdonságaik, amelyek eleve a digitális közeghez kötődnek, illetőleg analóg másolat előállításánál ezek a digitális közeghez kötődő sajátosságok torzulnak, megsemmisülnek, továbbá akár az is lehetséges, hogy nem is lehet analóg másolatot készíteni róluk. A legegyszerűbb born digital szövegfájl is több, mint a szöveges tartalmának nyomtatott változata: a fájlban tárolva vannak olyan metaadatok, amelyek az értelmezhetőség szempontjából fontos kontextuális információt tartalmazhatnak (pl. egy Word-dokumentum létrehozásának az időpontja).

Mindezek alapján a dHUpla rendszere eleve nem zár ki semmilyen formátumot, ugyanakkor kutathatóvá tenni, megjeleníteni, publikálni mindenekelőtt a szöveges tartalommal bíró digitális objektumok meghatározott körét tudja: szövegfájlokat, prezentációkat, e-maileket, egyéb szöveges üzeneteket (amennyiben szabványos formátumban állnak rendelkezésre), webes tartalmakat (blog-, weboldal-, közösségimédia- és egyéb exportokat, amennyiben szabványos formátumban állnak rendelkezésre). Ez utóbbiakkal kapcsolatban fontos újdonságot jelent, hogy a PIM-DBK rendszere alkalmas lesz a PIM gyűjtőkörébe vágó élő webes források szabályozott archiválására: kortárs írók, művészek által előállított, élő born digital tartalom befogadására és kezelésére, valamint publikálására és kutathatóvá tételére.

7 <https://digitalisjoletprogram.hu/hu/tartalom/kds-kozgyujtemenyi-digitalizalasi-strategia>, elérés: 2021. május 15.

8 Bánki Zsolt és Kómár Éva, szerk., *Fehér könyv. Módszertani útmutató a közgyűjteményi kulturális örökség digitalizálásához és közzétételéhez* (Emberi Erőforrások Minisztériuma, 2019), 123. https://ommik.hu/media/attachments/2019/12/09/fehr_knyv.pdf. Elérés: 2021. május 15.

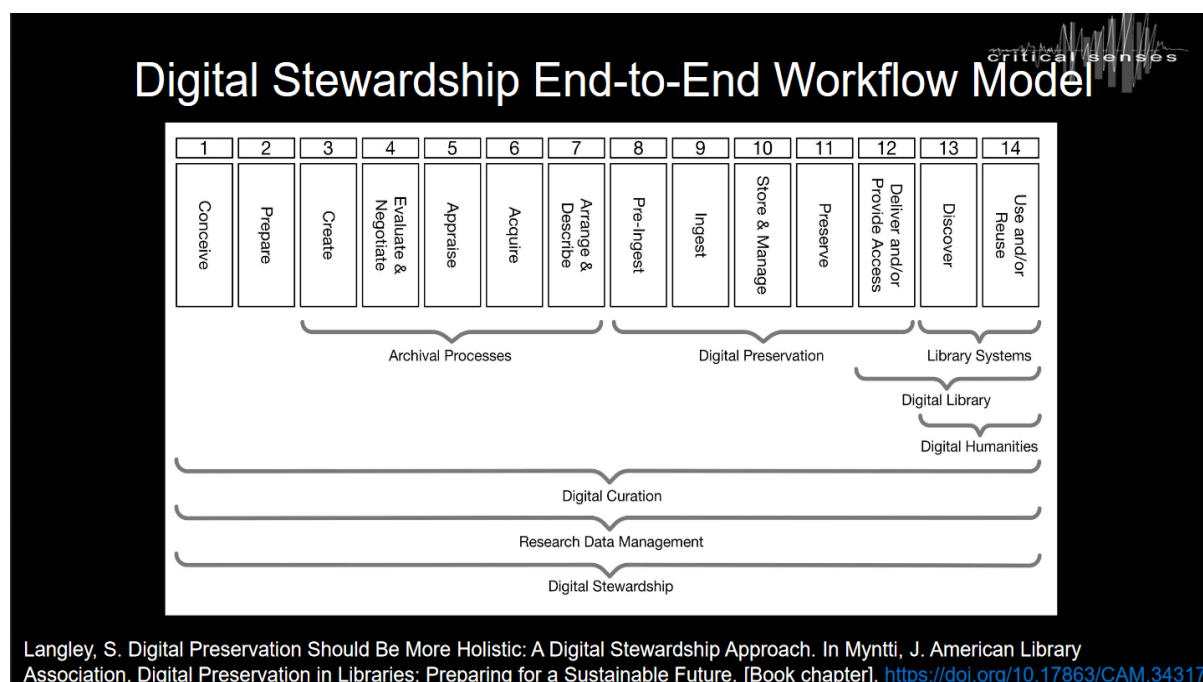
9 Uo., 123.



3. A PIM-DBK born digital workflow-jának kidolgozása előtt el kellett végeznünk a kutatómunkát: fel kellett tárnunk a vonatkozó szakirodalmat, valamint a hazai és nemzetközi jó gyakorlatokat. Bár a túlnyomórészt angol nyelvű szakirodalmi listánk hamar tekintélyesre duzzadt, összességében mégis inkább általában a digitális objektumok (és nem annyira speciálisan a born digital) kezelésére (digital curation v. digital stewardship) vonatkozó tételek dominálnak.¹⁰ A magyar nyelvű szakirodalomból kiemelendő a *Fehér könyv*, amely első változatában ugyan nem tartalmaz kifejezetten a born digital objektumok kezelésére vonatkozó részeket, a digitális megőrzésre vonatkozó általánosabb alapelvei viszont természetesen mérvadóak és követendőek. Számunkra alapvetően fontos volt a VII. fejezet (A digitális tartalmak menedzsmentje), amely többek között fontos terminológiai javaslatokat is tartalmaz. Ennek alapján használjuk pl. a digitális objektum terminus technicust.¹¹

3.1. Ugyancsak a *Fehér könyv*, valamint a nemzetközi szakirodalom, továbbá a tanulmányozott jó gyakorlatok alapján határoztuk meg az elméleti modelleket, amelyeket a born digital objektumok kezelésében is eredményesen lehet alkalmazni.

3.1.1. Bár a *Fehér könyv* a brit Digital Curation Centre Curation Lifecycle Modeljét ajánlja,¹² saját használatra egy bővebb modellt, Somaya Langley Digital Stewardship End-to-End Workflow Modeljét vettük alapul.¹³



Langley, S. Digital Preservation Should Be More Holistic: A Digital Stewardship Approach. In Myntti, J. American Library Association, Digital Preservation in Libraries: Preparing for a Sustainable Future. [Book chapter]. <https://doi.org/10.17863/CAM.34317>

1. ábra. Somaya Langley Digital Stewardship End-to-End Workflow Modelje

10 <https://www.zotero.org/groups/2532329/pim-dbk-dh-born-digital/library>. Elérés: 2021. május 15.

11 Bánki és Kómár, *Fehér könyv. Módszertani útmutató a közgyűjteményi kulturális örökség digitalizálásához és közzétételéhez*, 123.

12 Uo., 124. L. továbbá <https://www.dcc.ac.uk/guidance/curation-lifecycle-model>. Elérés: 2021. május 15.

13 Langley Somaya, „Digital Preservation Should Be More Holistic: A Digital Stewardship Approach” (Apollo - University of Cambridge Repository, 2018), <https://doi.org/10.17863/CAM.34317>.

Ez a modell a teljes digitális objektumkezelés folyamatát 14 fázisra bontja.¹⁴

Langley cikkében kiemeli, hogy a digitális objektumok kezelésében alkalmazott modellek hiányosak, nem terjednek ki a teljes életciklusra (amelynek fontos része pl. a digitális tartalom létrejöttének a megtervezése, vagy éppen a felhasználás, hasznosítás szempontjainak a figyelembevétele). Pedig a körültekintő tervezés segíthet a digitális megőrzés problémáinak (mennyiség, változatosság, elavulás, értelmezhetőség; l. fentebb) enyhítésében.

3.1.2. A digitális tartalmak menedzsmentjének, a megőrzés általános feladatainak a modellezésére világszerte a CCSDS által javasolt **Open Archival Information System (OAIS)** modell, valamint a hozzá kapcsolódó szabványcsalád a legelterjedtebb.¹⁵ Ez egy nemzetközi standard (ISO 14721:2012), amelyet a *Fehér könyv* is ajánl, valamint Magyarországon is van már példa az implementálására: a Magyar Nemzeti Levéltár born digital objektumkezelési gyakorlata is ezen alapul.¹⁶ A modell három fő ágenszt különböztet meg: a tartalom előállítója, kezelője és felhasználója. Az ágensek közötti adatcserére vonatkozó ajánlásokat fogalmaz meg: a digitális objektumokat ún. csomagokban javasolja kezelni, amelyekben kötelezően előírt elemek vannak. A SIP-csomag (Submission Information Package) tartalmazza az előállító által létrehozott tartalmat, valamint a létrehozás és a kezelőnek átadás körülményeit dokumentáló metaadatfájlokat. Ebből a kezelő AIP-csomagot (Archival Information Package) hoz létre, amely a SIP-csomagon kívül tartalmazza a megőrzési terv alapján előírt további kötelező elemeket: pl. a megőrzési formátumra konvertált reprezentációkat, azok metaadatait, mindenféle, a megőrzést megelőző művelet (pl. vírusellenőrzés, hashing stb.) dokumentációját, metaadatait. Az AIP-csomag célja kettős: egyrészt az archiválás, másrészt ebből jön létre a szolgáltatási csomag, a DIP (Dissemination Information

14 Nem a magyar terminológia egységesítésének, ill. bizonyos esetekben kialakításának a szándékával, hanem csupán a jelen cikk szóhasználatának bevezetése céljából közöljük ezen fázisok általunk használt magyar megfelelőjét:

- 1 Conceive: tervezés
- 2 Prepare: előkészítés
- 3 Create: létrehozás, előállítás
- 4 Evaluate & Negotiate: kiértékelés és megvitatás
- 5 Appraise: értékelés, besorolás
- 6 Acquire: beszerzés
- 7 Arrange & Describe: elrendezés és leírás
- 8 Pre-Ingest: a gyarapítás (átvétel) előkészítése
- 9 Ingest: gyarapítás (átvétel)
- 10 Store & Manage: tárolás és kezelés
- 11 Preserve: megőrzés
- 12 Deliver and/or Provide Access: szolgáltatás és/vagy hozzáférés biztosítása
- 13 Discover: feltárás
- 14 Use and/or Reuse: használat és/vagy újrahasznosítás.

15 CCSDS, *Reference Model for an Open Archival Information System (OAIS)*, *Magenta Book* (CCSDS - Consultative Committee for Space Data Systems, 2002), <http://public.ccsds.org/publications/archive/650x0b1.pdf>, Elérés: 2021. május 15. További szabványok: PAIMAS (<https://public.ccsds.org/Pubs/651x0m1.pdf>, elérés: 2021. május 15.); PAIS (<https://public.ccsds.org/Pubs/651x2g1.pdf>, elérés: 2021. május 15.)

16 <https://mnl.gov.hu/mnl/szkk/szabvanyok>, elérés: 2021. május 15.

Package), amelyet a felhasználók rendelkezésére lehet bocsátani. Ez a csomag az AIP bármely reprezentációját, valamint szükség esetén további konvertált reprezentációkat, azok dokumentációját, metaadatait tartalmazza.



2. ábra. Az Open Archival Information System (OAIS) alapfogalmai

A „csomagolás” az OAIS-modellben korántsem öncélú, számos haszna mellett két fontos mozzanat emelendő ki: a digitális objektumokra vonatkozó dokumentáció, ill. metaadataik együtt kezelhetők, mozgathatók magával az objektummal, továbbá azon digitális objektumok esetében, amelyek több különálló fájlból vagy akár könyvtárak struktúrájából állnak, lehetséges az objektum részeinek egyben tartása, valamint az eredeti felépítés megőrzése.

3.1.3. A csomagok tartalmára nézve is létrejöttek ajánlások, amelyek beépültek az OAIS-modellt alkalmazó szoftverekbe is, megkönnyítve a digitális kurátorok dolgát a csomagkészítés terén. Az egyik ilyen ajánlás a Library of Congress által kidolgozott BagIt, amely az Internet Engineering Task Force 8493-as ún. RFC-je (Request for Comments).¹⁷ A BagIt előírása alapján szükséges egy tetszőleges elnevezésű gyökérmappa (base directory). Ebben kötelezően 3 elem van: egy data elnevezésű mappa, ebben van a tartalom (payload). A tartalom tetszőlegesen összetett lehet (maga is egy vagy több mappa további almappákkal és fájlokkal stb.). Szükséges legalább egy manifest-<algorithm>.txt fájl, ahol az <algorithm> a fájlnevében helyettesítendő a használt checksumkészítési algoritmus nevével (pl. MD5). Ebben a fájlban benne kell

¹⁷ Kunze John A., Littman Justin, Madden Liz, Scancellia John és Adams Chris, *The BagIt File Packaging Format (V1.0)*, Request for Comments 8493 (RFC Editor, 2018), <https://doi.org/10.17487/RFC8493>. Request for Comments 8493 (RFC Editor, 2018) Elérés: 2021. május 15. Itt említendő még az elsősorban levéltári használatra kifejlesztett E-Ark sepcifikáció, amely a BagIt-nél jóval összetettebb csomagokat ír elő. <https://www.eark-project.com/>. Elérés: 2021. május 15.

lennie a csomag összes fájljára vonatkozó checksumértéknek. Szükséges végül egy bagit.txt fájl, amelyben két információt kell megadni: a Bagit csomagolási eljárás verzióját (jelenleg 1.0), és a TXT-fájlok karakterkódolását (pl. UTF-8, ez utóbbi az ajánlott). A csomag további címkefájlokat is tartalmazhat opcionálisan, amelyek elnevezésére és tartalmára, valamint a formátumára és kódolására a Bagit tesz javaslatokat. Pl. bag-info.txt ilyen adatokkal: Source-Organization, Contact-Name, Contact-Email stb. Lehetséges még címkefájlokat (metaadat-fájlokat) elhelyezni külön mappába is (tetszőleges mappanévvvel). Minden opcionális fájl checksumjának szerepelnie kell a manifestfájl(ok)ban. A Bagit csomagolóeljárást számos szoftver támogatja (Bitcurator, RODA).

4. A szakirodalom és a nemzetközi jó gyakorlatok alapján a fentebb kifejtett alapelveknek megfelelő workflow kidolgozásába kezdtünk a PIM-DBK-ban. Kiindulópontunk az volt, hogy a born digital objektumok kétféle csatornán kerülhetnek a rendszerünkbe: adathordozón, valamint hálózatosan (e-mailben elküldve, felhőtárhelyes hozzáférés megadásával stb.). Biztosítani kívánjuk továbbá, hogy a csatlakozó partnerek (akár magányszemélyek, akár más közgyűjtemények vagy egyéb szervezetek) képesek legyenek az általunk fejlesztett rendszer segítségével a saját born digital tartalmaik kezelésére. Rendszerünket tehát úgy kellett megtervezni, hogy mind a PIM belső, már meglévő gyűjteményi elemeit (akár adathordozón, akár szerveren tárolt anyagról van szó), mind a kívülről (akár adathordozón, akár hálózaton) érkező anyagokat fel tudjuk dolgozni.

4.1. Az adathordozón tárolt anyagok kezelése speciális hardveres és szoftveres infrastruktúrát igényel. Ennek oka az, hogy az adathordozón található ideiglenes vagy akár törölt fájlok, magának az adathordozónak a fájlstruktúrája stb. fontos, megőrzendő kontextuális információt jelentenek, továbbá szükséges biztosítani azt is, hogy a tartalom másolása szakszerűen, az fájlok integritásának megőrzésével történjen. Ebből következően az adathordozók gyűjteményi archiválásához szükség van megfelelő számítógépre, amely fel van szerelve az ún. digitális törvényszéki módszerek alkalmazásához elengedhetetlen hardveres és szoftveres kiegészítőkkal (írásblokkoló, lemezképkészítő, fájlvalidáló, vírus- és malware-ellenőrző stb.). A kialakult nemzetközi standardoknak megfelelően a PIM-DBK is az USA-beli Digital Intelligence termékét, a FRED-et (Forensic Recovery of Evidence Device),¹⁸ vagy azzal ekvivalens nagy teljesítményű PC-t kíván szolgálatba állítani, amelyre telepíteni kell a Bitcurator nevű Linux-disztribúciót.¹⁹ A FRED és a Bitcurator segítségével minden olyan feladat megoldható, amely a born digital tartalmak gyűjteményezéséhez elengedhetetlen. Az adathordozó tartalmáról biztonságosan bitszintű másolat készülhet, a tartalom szűrhető (eltávolíthatók az archiválni nem kívánt vagy az archiválásból jogilag kizárt fájlok, akár a donor közreműködésével is) stb., továbbá minden műveletről részletes és szabványos dokumentáció készülhet.

4.2. A nem adathordozón tárolt tartalom esetében is szükséges a Bitcurator használata, hiszen a born digital objektumok gyakran tartalmaznak kiszűrendő érzékeny információkat (pl. védett személyes adatokat, a GDPR által a felhasználáshoz hozzájárulást előíró

¹⁸ <https://digitalintelligence.com/products/fred/>, elérés: 2021. május 15.

¹⁹ <https://bitcurator.net/>, elérés: 2021. május 15.



adatokat). A Bitcurator megfelelő konfigurációval jelentősen megkönnyíti ezek kezelését, továbbá tartalmaz egy Bagger nevű eszközt is, amellyel (akár adathordozóról származik a tartalom, akár online forrásból) elvégezhető a Baglt specifikációnak megfelelő csomagolás, azaz előállhat a gyűjteményezési folyamat bemeneteként szolgáló SIP-csomag. Ez a csomag kerül be a PIM-DBK által fejlesztett digitális bölcsészeti platform (dHUpla) rendszerébe.

4.3. A SIP-csomag tartalma igen változatos lehet, a data mappa (l. fentebb) egyetlen fájl, de akár sokszintű könyvtárstruktúrát is tartalmazhat. A leglényegesebb alapelv, hogy a SIP-csomagnak tartalmaznia kell az eredeti (beadási) formátumot, továbbá minden dokumentációt, amely a bekerüléssel kapcsolatos (a donorral kötött szerződéseket, a Bitcurator által előállított jelentéseket stb.).

5. A born digital tartalmak a dHUpla rendszerében. A PIM-DBK digitális bölcsészeti platformja a born digital tartalmak kezelésére egy OAIS-kompatibilis teljes körű repozitóriumszolgáltatást nyújtó nyílt forráskódú szoftvert, a portugál Keep Solutions RODA nevű termékét²⁰ szándékozik használni. Ez a szoftver képes befogadni az előállt SIP-csomagokat, amelyekből tetszőleges tartalmú AIP-csomagokat állít elő. A csomagokkal minden, a repozitóriumokban megszokott művelet elvégezhető (akár API-n keresztül, automatizáltan és tömegesen is): fájlok átcsoportosítása, könyvtárstruktúra kialakítása, megváltoztatása, a fájlok törlése, betekintés stb. Mivel számos nyílt forráskódú eszközt magába integrál, a RODA képes az AIP-csomagok kialakításához műveleteket elvégezni: fájlkonverziók, (újra)hashelés, validálás stb. Hosszútávú megőrzésre szabványos formátumra konvertáljuk a bemeneti formátumokat, az AIP-csomagok az eredeti mellett ezeket is tartalmazzák. A dHUpla rendszere ezeknek a szabványos formátumoknak a megjeleníthetőségét, kutathatóságát is biztosítja.

A RODA kimenetként ugyancsak tetszőlegesen összeállított DIP-csomag letöltését is lehetővé teszi. Azon born digital tartalmak esetében, amelyek tudományos szempontból kitüntetett figyelmet érdemelnek (írói kéziratok, feljegyzések, levelezések stb.), a dHUpla forráskiadási rendszere²¹ tudományos kiadást állít elő, amely külön felületen megtekinthető és kutatható, valamint nem utolsósorban olyan szemantikus annotációt kap, amely lehetővé teszi különféle szemantikus kapcsolatok létrehozását (pl. jelentős nemzetközi adatbázisokkal vagy éppen a Nemzeti Névtérrel).

²⁰ <https://www.keep.pt/en/producs/roda-long-term-digital-preservation-repository-solution/>, elérés: 2021. május 15.

²¹ A forráskiadás rendszere külön részrendszer a dHUplán belül, amelynek működése nem képezi jelen előadás tárgyát.

Bibliográfia

- Bánki, Zsolt és Kómár, Éva, szerk. *Fehér könyv. Módszertani útmutató a közgyűjteményi kulturális örökség digitalizálásához és közzétételéhez*. Emberi Erőforrások Minisztériuma, 2019. https://ommik.hu/media/attachments/2019/12/09/fehr_knyv.pdf.
- CCSDS. *Reference Model for an Open Archival Information System (OAIS)*, Magenta Book. CCSDS - Consultative Committee for Space Data Systems, 2002. <http://public.ccsds.org/publications/archive/650x0b1.pdf>.
- Magyar Közlöny. „34/2016. (XI. 30.) EMMI rendelet az elektronikus formában tárolt iratok közlevéltári átvételének eljárásrendjéről és műszaki követelményeiről”. Elérés 2021. június 15. <https://magyarkozlony.hu/dokumentumok/afc7a7bc8a8b8c9e22db85179ec5932bc5fa05e9/megtekintes>.
- Kunze, John A., Littman, Justin, Madden, Liz, Scancella, John és Adams, Chris. *The BagIt File Packaging Format (V1.0)*. Request for Comments 8493. RFC Editor, 2018. <https://doi.org/10.17487/RFC8493>.
- Langley, Somaya. „Digital Preservation Should Be More Holistic: A Digital Stewardship Approach”. Apollo - University of Cambridge Repository, 2018. <https://doi.org/10.17863/CAM.34317>.