

Kriteriji za odabir formata za dugoročnu pohranu elektroničkoga gradiva

Ivanjko, Tomislav; Lončarić, Ksenija

Source / Izvornik: **Arhivi i domovinski rat, 2016, 413 - 424**

Conference paper / Rad u zborniku

Publication status / Verzija rada: **Submitted version / Rukopis poslan na recenzijski postupak (preprint)**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:131:142435>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom](#).

Download date / Datum preuzimanja: **2024-10-04**



Sveučilište u Zagrebu
Filozofski fakultet
University of Zagreb
Faculty of Humanities
and Social Sciences

Repository / Repozitorij:

[ODRAZ - open repository of the University of Zagreb
Faculty of Humanities and Social Sciences](#)



Kriteriji za odabir formata za dugoročnu pohranu elektroničkoga gradiva

Dugoročna pohrana elektroničkoga gradiva je jedan od glavnih zadataka arhivske prakse. Razvoj tehnologije omogućio je kvalitetnije rukovanje gradivom, ali i postavio pitanja o poteškoćama na koje se može naići u procesu pohrane. Jedno od tih pitanja tiče se formatâ elektroničke pohrane čiji je glavni cilj očuvanje integriteta, dostupnosti i iskoristivosti gradiva. Ovaj rad bavit će se definiranjem i raspravom o kriterijima za odabir formata za dugoročnu pohranu.

Ključne riječi: format zapisa, dugoročna pohrana, elektroničko gradivo, kriteriji odabira baštinske institucije

1. Uvod

Arhivi, knjižnice i muzeji kao baštinske ustanove glavni su akteri u prikupljanju, pohrani, osiguravanju pristupa i očuvanju zapisa od trajna značaja za stvaratelje, imatelje i krajnje korisnike. Arhivi kao „ustanove specijalizirane za čuvanje, zaštitu, obradu i korištenje izvorne pisane baštine, čuvaju autentičan i vjerodostojan trag prošlosti i zbivanja koje ti zapisi dokumentiraju“¹. U svojim svakodnevnim aktivnostima ove ustanove rukuju velikom količinom gradiva, a razvojem tehnologije elektroničko gradivo dobiva primat nad ostalim oblicima. Odabir, a tako i vrednovanje, gradiva za dugoročnu pohranu nužan je postupak kako bi se očuvalo ono što je zaista potrebno i važno za stvaratelja, imatelja te u konačnici i korisnika. Jednaku ili veću važnost ima odabir formata zapisa za dugoročnu pohranu elektroničkoga gradiva. Kriteriji odabira formata fokusiraju se na mogućnost korištenja zapisa neovisno o trenutnom razvoju i dostupnosti tehnologije, na pitanje intelektualnog vlasništva i prava, postojanje podrške i očuvanje vjerodostojnosti. Odabir formata u kojemu će zapis nastati te se

¹ Hrvatski državni arhiv. Arhivska služba. 2013. URL: <http://www.arhiv.hr/arhiv2/Arhivskasluzba/index.htm> (18.7.2016.)

prvotno pohraniti određuje i vrstu i način kasnije dugoročne pohrane. O tome ovisi i mogućnost pristupa elektroničkom gradivu. Radi osiguranja kasnije mogućnosti pristupa, ustanove nadležne za pohranu gradiva i građe i javne ustanove koje stvaraju i čuvaju gradivo trebale bi provesti analizu rizika korištenja pojedinih formata čime bi osigurale korištenje formata zapisa s najboljim karakteristikama i koji zadovoljava najviše kriterija. Logistička i financijska ograničenost često onemogućuju provođenje analize rizika na licu mjesta. U tom slučaju zadatak se prepušta nadležnim stručnjacima koji će utvrditi koji format zapisa za koje potrebe i s kojim medijima najbolje surađuje.

Velika količina gradiva koja pristiže u arhive rijetko dolazi u jednom, standardiziranom obliku. Od različitih struktura i organizacija sadržaja unutar dokumenata, preko različitih medija do stanja očuvanosti, arhivsko gradivo zahtijeva različite načine pristupanja i odnošenja s njime. Kada je riječ o elektroničkom gradivu, velike razlike u formatima i medijima pohrane mogu predstavljati problem. Stvaratelji gradiva prilikom stvaranja zapisa prilagođavaju odabir formata svojim potrebama i navikama. Svaki zaposlenik jedne tvrtke, ustanove, agencije, službe može i često ima različite načine zapisivanja, strukturiranja i organizacije, a time i različito znanje i shvaćanje važnosti dugoročne pohrane. U takvim prilikama nastaju različiti zapisi i dokumenti, isti podaci mogu biti spremljeni u različitim formatima, mogu se izgubiti i u najgorem slučaju izbrisati. No, kada treba započeti proces dugoročne pohrane koji zahtijeva pomno planiranje, utvrđivanje najboljeg formata nužno za sobom vodi i opsežan proces migracije zapisa koji nisu u skladu s njime. Osim financijskih troškova, vremensko ograničenje i čest manjak stručnog osoblja onemogućuje ili znatno usporava sam projekt. U takvim slučajevima, kao što je već bilo napomenuto, važnu ulogu imaju provedena istraživanja i studije slučaja na temu preporučenih formata zapisa za dugoročnu pohranu.

Cilj ovoga rada je prikazati kriterije odabira (engl. *selection criteria*) formata zapisa za dugoročnu pohranu elektroničkoga gradiva. Na temelju sekundarne analize identificirane postojeće literature, autori predstavljaju problematiku definiranja i usklađivanja definicija termina formata zapisa, podjele i kategorizaciju formata zapisa. Drugi dio rada usmjeren je prema predstavljanju različitih pristupa odabiru formata zapisa te nabraja i pojašnjava kriterije odabira s ciljem definiranja smjernica koje će pomoći institucijama u njihovom poslovanju i odnošenju s gradivom.

2. Glavne odrednice, definicije i koncepti

Jedan od temeljnih problema koji je kao takav prepoznat od strane arhivističke zajednice je nedostatak standardizirane definicije i shvaćanja formata, a zatim i postojanje te upotreba različitih formata i njihovog kodiranja. Usmjerenje prema određenoj definiciji termina i prihvaćanje pristupa ovisi o izvornoj znanstvenoj zajednici i njenim preporukama. Vrsta gradiva koja se stvara tokom vremena razlikuje se ovisno o struci aktera i time je pitanje određenja pojmova riješeno unutar okvira njihove struke. Kako bi mogli uspoređivati formate i provesti valjanu analizu, potrebno je uniformno određenje glavnih pojmova, koncepata i podjela. Ovo poglavlje je pokušaj usklađivanja definicija iz različitih izvora s ciljem dolaženja do definicije koja će uspjeti uokviriti ono najbitnije. Krenuvši od formata zapisa, fokus će biti prebačen na različite vrste i podjele formata zapisa za dugoročnu pohranu elektroničkoga gradiva.

2.1 Format zapisa

Arhivistički rječnik definira format zapisa (engl. *file format*) kao „oblik u kojemu su podatci u datoteci kodirani“². Proširenu definiciju daje InterPARES 2 te navodi kako je format zapisa „organizacija podataka u digitalnim objektima, stvorena kako bi se olakšala pohrana, dohvat, obrada, predstavljanje i/ili prijenos podataka pomoću softvera“³. Sljedeća definicija navedena u završnom izvješću InterPARES 2 projekta preuzeta je iz „The PREMIS *Data Dictionary for Preservation Metadata*“ koji definira format zapisa kao „specifičnu, prethodno ustanovljenu strukturu za organizaciju digitalnog zapisa ili toka podataka“⁴.

Svaka od ovih definicija dotiče neki od aspekata i svojstava formata. Budući da se ovaj rad bavi problemom odabira formata, njihovim svojstvima i poteškoćama na koje nailaze arhivisti, bibliotekari i muzeolozi u izvršavanju svoje prakse, potrebno je proširiti postojeće definicije i stvoriti novu koja će navedeno obuhvatiti u jednostavnu i smislenu cjelinu. U ovom radu format zapisa predstavlja prethodno ustanovljenu Ostrukturu podataka digitalnih objekata

² Mihaljević, M., Mihaljević, M., Stančić, H. Arhivistički rječnik. Zagreb: Zavod za informacijske studije. 2015.

³ The InterPARES 2 Project. General Study 11 Final Report: Selecting Digital File Formats for Long-Term Preservation, 2007. str. 1.

⁴ PREMIS Working group, Data Dictionary For Preservation Metadata : OCLC Online Computer Library Center, Inc. i RLG. 2005.

koja ovisi o vrsti zapisa te je stvorena radi olakšanja pohrane, dohvata, obrade, predstavljanja i/ili prijenosa podataka pomoću softvera. Ova definicija poslužit će kao temelj za određivanje i vrednovanje kriterija odabira formata zapisa koji će kasnije u radu biti obrazloženi.

2.2. Kategorizacija formata zapisa

Formati zapisa međusobno se razlikuju po svojim specifikacijama, dostupnosti i inačici. Specifikacije formata određuju mogućnost njegove uporabe, namjenu i eventualni uvid u kodiranje formata. Nisu svi formati javno i slobodno dostupni i time specifikacije nisu otvoreno dostupne svima koji njima žele pristupiti. UK Data Archive formate zapisa dijeli na **zaštićene** (u smislu autorskih prava) (engl. *proprietary file formats*), **dostupne** (engl. *available file formats*) i **otvorene formate** (engl. *open file formats*)⁵. InterPARES 2 rječnik⁶ zaštićenim formatima naziva one koji su privatno posjedovani i licencirani od neke tvrtke, institucije ili organizacije. Njihove specifikacije nisu javno dostupne kako bi se spriječilo krivotvorenje ili kopiranje/dupliciranje formata kao proizvoda. Za razliku od zaštićenih, otvoreni formati zapisa javno su dostupni i dozvoljeno je njihovo mijenjanje, prilagođavanje i korištenje formata bez ograničenja ili troškova. Treći tip formata zapisa spoj je prijašnja dva. Dostupni formati su zaštićeni formati s javno dostupnim specifikacijama. InterPARES 2 projekt u svome završnom izvješću naglašava kako pojam otvorenog formata nema jednoznačnu definiciju. S jedne strane, njegove specifikacije su dostupne za korištenje, a stvaraju se i koriste pomoću softvera s otvorenim kodom (engl. *open source software*). Takav tip softvera razvija se i poboljšava putem javne suradnje i besplatno se dijeli uz zahtjev da se sljedeće inačice tog formata također besplatno dijele s ciljem poticanja daljnjeg razvoja⁷. S druge strane, otvoreni formati mogu biti oni koji nisu proizvod privatne organizacije i također, kao u prvome značenju, moraju biti svima dostupni za rukovanje s njima⁸. Problem koji se javlja kod zaštićenih formata odnosi se na njihovu dostupnost u određenom periodu. Vlasnik nekog formata zapisa u bilo kojem trenutku može odlučiti promijeniti ili zabraniti upotrebu tog formata zapisa što utječe na gradivo koje

⁵ The InterPARES 2 Project. General Study 11 Final Report: Selecting Digital File Formats for Long-Term Preservation, 2007. str. 3.

⁶ The InterPARES 2 Project. The InterPARES 2 Project Dictionary. 2016.

⁷ Mihaljević, M., Mihaljević, M., Stančić, H. Arhivistički rječnik. Zagreb: Zavod za informacijske studije. 2015. str. 138.

⁸ The InterPARES 2 Project. General Study 11 Final Report: Selecting Digital File Formats for Long-Term Preservation. 2007. str. 3.

arhiv ili druga baštinska institucija posjeduje i prisiljava ih na migraciju. Također, promjene u licenci formata zapisa mogu uvjetovati nove troškove i ograničenja glede upotrebe softvera za pristup i korištenje gradiva.

Sljedeća podjela formata zapisa je na **formate za očuvanje** (engl. *preservation formats*) i **formate za pristup** (engl. *access format*). Formati za očuvanje koriste se i prilagođeni su za dugoročnu pohranu elektroničkih zapisa u digitalnom arhivu, dok su formati za pristup prilagođeni za korištenje od strane korisnika u trenutku pristupa tome zapisu⁹. Funkcija i smisao dugoročne pohrane obuhvaća obje vrste formata. Osiguranje pristupa mora biti omogućeno kako bi proces dugoročne pohrane imao smisla. Glavni cilj, u konačnici, je pristup gradivu čime dolazimo do temeljne povezanosti između formata očuvanja i formata pristupa u kontekstu postizanja i održavanja procesa dugoročne pohrane. U određivanju kriterija odabira formata u kojemu će zapisi biti dugoročno pohranjeni, potrebno je uzeti u obzir mogućnost pristupa pojedinom formatu koja ovisi o dostupnoj tehnologiji što zaokružuje dosad navedeno. Podjela formata zapisa za dugoročnu pohranu prema dostupnosti i raširenosti formata je ključna za proces i planiranje dugoročne pohrane. U tom kontekstu važno je odrediti koja je vrsta formata najisplativija i koja zadovoljava sve ili najveći broj kriterija. Kao što je već spomenuto, baštinske ustanove, a ponajviše arhivi, preuzimaju gradivo od stvaratelja u raznim oblicima, stanjima i na različitim medijima. Zaštićeni formati zapisa ne odgovaraju potrebama tih institucija u projektima očuvanja jer njihovo korištenje donosi i velike troškove. Budući da se najčešće mogu koristiti isključivo s programima i aplikacijama osobe ili institucije koja na njih polaže autorska prava, njihovo korištenje u budućnosti ovisi o uspjehu i održivosti proizvoda te organizacije. Financijske mogućnosti ili, češće, financijska ograničenja arhiva sužuju izbor formata na one koji će pružiti najviše, a zahtijevati najmanje. To se prvenstveno odnosi na zahtjeve koje će u budućnosti imati korisnici, a moguće ih je predvidjeti u procesu planiranja i izgradnje sustava.

Formate zapisa moguće je podijeliti također po kriteriju preporučenosti i prihvaćenosti od strane različitih institucija, organizacija ili tijela. Prema Nizozemskom institutu za znanstvene informacijske usluge (engl. *The Netherlands Institute for Scientific Information*

⁹ Barnes, I. Preservation of Word Processing Documents : Australian Partnership for Sustainable Repositories (APSR). 2006. str. 4

Service (NISIS) standardni formati zapisa su „široko dostupni i prihvaćeni, izrazito su interoperabilni, ne zahtijevaju kompresiju podataka i podržavaju očuvanje metapodataka“¹⁰. Pojam standardnog formata zapisa ne označava samo jedan koji se koristi i preporuča u praksi očuvanja. Za svaku vrstu digitalnog objekta postoji format koji se preporuča od strane standardizacijskih organizacija poput Međunarodne organizacije za normizaciju (engl. zove International Organization for Standardisation (ISO)) ili Imaging Industry Association (I3A). Jedno shvaćanje standardnog formata zapisa orijentira se prema tim formatima kao jedinama koji mogu imati taj status. Za pohranu slika ISO preporuča nekoliko standardnih formata zapisa: TIFF/IT (ISO 12639:2004), PNG (ISO/IEC 15948:2004) i JPEG 2000 (ISO/IEC 15444:2003, 2004, 2005, 2007, 2008). Primjer za preporuku formata za pohranu dokumenata možemo također naći kod ISO standarada: PDF/A (ISO Standard 19005-1. *Document File Format for Long-Term Preservation*). PDF/A je primjer zaštićenog formata koji unatoč preporuci i dalje nije u potpunosti prihvaćen kao format za dugoročnu pohranu¹¹. Drugo shvaćanje standardnog formata zapisa ima Library and Archives of Canada unutar okvira „de facto standardnih“ formata koji su „iako zaštićeni, priznati kao formati i tipovi zapisa koji su postali industrijski standardi zbog svoje sveprisutnosti i podrške, a ne zato jer su službeno odobreni od strane standardizacijske organizacije“¹². Prema The Public Records Office of Victoria, Australija (PROV) preporuča se izbor industrijskog standardnog formata (engl. *industry standard format*) u slučajevima kada nezaštićeni formati (engl. *non-proprietary formats*) s javno dostupnim specifikacijama nisu dostupni. Iako su česte preporuke da zaštićene formate zapisa treba izbjegavati i okrenuti se prema otvorenim formatima, PROV suprotno napominje kako je smisao industrijskih standardnih formata njihova uloga u ekonomiji te kako tržište diktira upotrebu pojedinih formata u određenom periodu. Elektronički zapisi pohranjeni u nekom od zaštićenih formata zapisa koji dominiraju tržištem neće zahtijevati učestale migracije i povoljni su za dugoročnu pohranu. Ova teza kreće od pretpostavke da tržište upravlja industrijskim

¹⁰ Park, E. G., Oh, S. Examining Attributes of Open Standard File Formats for Long-term Preservation and Open Access. *Information Technology and Libraries*, 31, 4(2012), str. 46.

¹¹ Park, E. G., Oh, S. Examining Attributes of Open Standard File Formats for Long-term Preservation and Open Access. *Information Technology and Libraries*, 31, 4(2012), str. 46.

¹² The InterPARES 2 Project. General Study 11 Final Report: Selecting Digital File Formats for Long-Term Preservation. 2007.

standardima putem zakona tržišta. Proizvod s najvećim udjelom na tržištu vrlo teško gubi svoju dominantnu poziciju¹³.

„Stabilni“ (engl. *stable*) formati zapisa određeni su svojom glavnom karakteristikom: stabilnošću. To obilježje se odnosi na njihovu kompatibilnost s prethodnim inačicama formata zapisa, a često se koristi od strane softverske industrije¹⁴. Stabilnost formata zapisa u ovome pogledu trebala bi biti ključna karakteristika i kriterij odabira formata radi smanjenja obima posla i u konačnici troškova projekata dugoročne pohrane. Kompatibilnost s prethodnim inačicama osigurava arhivima dovoljno vremena da uvide promjene koje se događaju i da se uspiju njima prilagoditi u pogledu očuvanja gradiva.

3. Kriteriji za odabir formata za dugoročnu pohranu

U procesu dugoročne pohrane utvrđivanje odgovarajućeg formata jedan je od osnovnih preduvjeta za uspješno razvijanje modela i politika dugoročne pohrane različitih vrsta građe. Svaka ustanova koja se bavi dugoročnom pohranom i očuvanjem elektroničkog gradiva kroz razvijanje repozitorija susreće se u postupku s pitanjem prihvaćanja ili odbacivanja pojedinog formata kao prihvatljivog za dugoročnu pohranu. Pitanje odabira prihvatljivih formata neizbježno otvara raspravu o kriterijima odabira te time otvara raspravu o nizu tema vezanih kako uz kreiranje modela i politika dugoročnog očuvanja na razini ustanove, tako i na praktičnoj razini razvijanja softverskih rješenja koja će se koristiti kod izrade samog repozitorija.

U nastavku rada donosi se pregled relevantnih istraživanja koja su se bavila definiranjem kriterija za odabir formata. Pregledom istraživanja identificiran je značajan broj osobina koje se navode kao primarni kriteriji kod odabira formata za dugoročnu pohranu koji se razlikuju u pristupu i odabiru kriterija.

Becker i Rauber u svome radu¹⁵ koji se bavi analizom kriterija kod digitalne pohrane centralno mjesto daju odabiru pravilnog formata dugoročne pohrane: „...značajan dio rizika kod

¹³ The InterPARES 2 Project. General Study 11 Final Report: Selecting Digital File Formats for Long-Term Preservation. 2007. str. 4

¹⁴ The InterPARES 2 Project. General Study 11 Final Report: Selecting Digital File Formats for Long-Term Preservation. 2007. str. 5

¹⁵ Becker, C., Rauber, A. Decision Criteria in Digital Preservation : What to Measure and How. Journal of the American Society for Information Science and Technology (JASIST), 6, 62, 6(2011). str. 1009-1028.

digitalnih sadržaja odnosi se na format „...što je često ključni kriterij odabira.“ Dalje navode nekoliko ključnih kriterija poput važnosti standardizacije formata (npr. putem ISO norme), njegove kompleksnosti ili otvorenosti, podrške u zajednici te kompatibilnosti sa standardnim preglednicima. Sullivan¹⁶ daje popis poželjnih atributa formata za dugoročnu pohranu iz arhivske perspektive navodeći kako format mora biti neovisan o uređaju, samodostatan, samo-opisujući, transparentan, pristupačan i lako dostupan. Rauch, Krottmaier i Tochtermann¹⁷ navode kriterije koji se sastoje od tehničkih karakteristika (npr. otvorena specifikacija, kompatibilnost i standardizacija) i tržišne pozicije formata (npr. trajanje jamstva, trajanje podrške, tržišnog udjela te broja nezavisnih proizvođača formata). Sahu¹⁸ se usredotočuje na kriterije razvijene od strane Britanskog državnog arhiva koji uključuju otvoreni standard, sveprisutnost, stabilnost, podršku za metapodatke, funkcionalnost, interoperabilnost i održivost.

Christensen prema Barve¹⁹ navodi pet osnovnih kriterija za odabir formata za dugoročnu pohranu:

- 1) format treba biti jednostavan za opis, razumijevanje i implementaciju,
- 2) format ne smije ovisiti o specifičnom hardveru za korištenje,
- 3) format ne smije ovisiti o specifičnom operativnom sustavu,
- 4) format ne smije ovisiti o patentiranom softveru i
- 5) format mora biti robustan i otporan na moguće kvarove.

Nacionalna knjižnica Nizozemske (*Koninklijke Bibliotheek*) u svojoj analizi formata za dugoročno očuvanje²⁰ navodi sedam osnovnih kriterija:

- 1) **Otvorenost** (engl. *openess*) – uključuje standardizaciju formata, moguća ograničenja korištenja i mogućnosti čitanja formata slobodno dostupnim alatima. Kriterij se odnosi

¹⁶ Sullivan, S. J. An Archival/records Management Perspective on PDF/A. *Records Management Journal*, 16, 1(2006). str. 51-56.

¹⁷ Rauch, C., Krottmaier, H., Tochtermann, K. File-Formats for Preservation: Evaluating the Long-Term Stability of File Formats. *Proceedings ELPUB2007 Conference on Electronic Publishing* (Beč). 2007.

¹⁸ Sahu, D. K. Long Term Preservation: Which File Format to Use. *Workshops on Open Access & Institutional Repository* (Chennai). 2004.

¹⁹ Barve, S., File Formats in Digital Preservation. *International Conference on Digital Libraries (ICSD)* (Bangalore) 2007. str. 239-248.

²⁰ Rog, J. & Van Wijk, C. *Evaluating File Formats for Long-Term Preservation*, The Hague: National Library of the Netherlands. 2008. str. 83-90.

na relativnu lakoću prikupljanja informacija o strukturi samog formata, gdje se veća razina znanja o formatu smatra boljom podlogom za uspješno dugoročno očuvanje.

- 2) **Prihvaćenost** (engl. *adoption*) – uključuje kriterije razine korištenja na svjetskoj razini kao i u svrhu arhivskog formata unutar baštinskog sektora. Kriterij se odnosi na popularnost i sveprisutnost formata u zajednici *developer*a koji će razvijati alate za čitanje pojedinih formata, kao i na prisutnost i primjenu u projektima očuvanja institucija koje se bave dugoročnim očuvanjem.
- 3) **Kompleksnost** (engl. *complexity*) – uključuje tri karakteristike: čitljivost ljudima, stupanj kompresije i raznovrsnost. Kriterij indicira koliko je složen postupak dešifriranja samog formata. Ako je format veoma složen i postoji opasnost da se ne razumije u potpunosti, to može predstavljati problem kod dugoročne pohrane i pristupa.
- 4) **Mehanizmi tehničke zaštite** (engl. *technical protection mechanism (DRM)*) – Uključuje karakteristike zaštite lozinki, zaštite kopiranja, digitalnih potpisa, zaštite od ispisa te zaštite od ekstrakcije sadržaja (engl. *content extraction protection*). Ovaj kriterij odnosi se na mogućnosti formata da ograniči pristup sadržaju. Takav ograničeni pristup sadržaju može predstavljati problem kod migracije sadržaja gdje je potreban stalan pristup digitalnom objektu.
- 5) **Samo-dokumentiranost** (engl. *self-documentation*) – Kriterij je razrađen na dvije razine – razina metapodataka te razina tehničkog opisa. Kriterij uključuje mogućnosti formata da u sebe ugradi metapodatke vezane ili uz sam objekt ili uz specifičnosti samog formata. Kada format omogućuje ugradnju metapodataka s podacima o nekom resursu (npr. ime autora, opis, itd.) ili podacima o specifičnosti samog formata u zaglavlje (npr. kako pročitati zapis), tada format podržava očuvanje podataka bez upućivanja na vanjske izvore. Što se više zna o digitalnom objektu u nekom formatu, to će ga se bolje moći razumjeti i koristiti u budućnosti.
- 6) **Robusnost** (engl. *robustness*) – Kriterij se odnosi na otpornost formata na kvarove u strukturi formata (engl. *corruption*) i česte promjene verzija i generacija specifikacije samog formata. Format koji se često mijenja i nije stabilan može ugroziti pristup resursima kod dugoročnog očuvanja.

- 7) **Zavisnost** (engl. *dependencies*) – Kriterij ukazuje na ovisnost formata o specifičnim okruženjima ili drugim vanjskim resursima (npr. specifični fontovi ili postupak dekodiranja). Visoka ovisnost formata o vanjskim faktorima predstavlja rizik kod dugoročnog očuvanja, iz razloga što čitljivost formata ovisi o faktorima koji nisu kontrolirani od strane proizvođača formata.

Komplementarnu analizu koja je željela ispitati prikladnost pojedinih formata u svjetlu dugoročnog očuvanja, provela je i Britanska knjižnica gdje je definirano devet kriterija za odabir formata na temelju kojih je provedena analiza:

1. **Prihvaćenost i korištenje** (engl. *adoption and usage*) – odnosi se na općenitu razinu korištenja i prihvaćenosti formata, kako na općenitoj razini, tako i unutar baštinskih ustanova koje imaju iskustvo rada s određenim formatima.
2. **Programska podrška** – odnosi se na ukupnu programsku podršku i razvojnu zajednicu koja daje podršku i razvija određeni format.
3. **Dostupnost dokumentacije i uputa** – odnosi se na dostupnost praktičnih uputa i dokumentacije s preporukama i postupcima u rješavanju problema u procesu implementacije.
4. **Kompleksnost** – odnosi se na kompleksnost formata u smislu upravljanja sadržajima koji su zapisani u tom formatu te razinu stručnosti koja je potrebna kako bi se formatom moglo pouzdano upravljati.
5. **Uključeni ili pridodani sadržaj** (engl. *embedded or attached content*) – odnosi se na potencijal formata da u sebe uklopi ili pridoda sadržaje sličnih formata i mogućih posljedica takve prakse.
6. **Vanjske zavisnosti** (engl. *external dependencies*) – odnosi se na ovisnost formata o specifičnim okruženjima ili drugim vanjskim resursima bez kojih sam format ne može funkcionirati.
7. **Pravni okvir** (engl. *legal issues*) – odnosi se na pravne implikacije korištenja, upravljanja ili očuvanja pojedinog formata.
8. **Mehanizmi tehničke zaštite** (engl. *technical protection mechanism (DRM)*) – odnosi se na mehanizme koji ograničavaju korištenje ili upravljanje pojedinim formatima.

9. **Ostali rizici** – odnosi se na ostale rizike koji nisu obuhvaćeni prethodnim kategorijama ili se ne mogu uklopiti u neku od kategorija, a susreću se u praktičnim situacijama kod rada s formatima.

Kao što je vidljivo iz pregleda literature, odabir formata za dugoročnu pohranu već se na prvom koraku odabira kriterija susreće s nizom različitih pristupa i odabira. Jedna od studija koja je pokušala identificirati ključne karakteristike koje treba tražiti u formatu koji želimo koristiti u svrhu dugoročnog očuvanja ili stvaranja preporuka ili politika oko odabira prikladnih formata za dugoročnu pohranu, jest i InterPARES 2 istraživanje²¹ koje je provelo komparativnu studiju 20 projekata koja se bave dugoročnim očuvanjem gradiva te na temelju analize donosi pet glavnih kriterija kojima su se navedeni projekti vodili kod odabira formata:

- 1) format je u širokoj primjeni (engl. *widely adopted*),
- 2) vlasništvo nad formatom nije zaštićeno (npr. patentom) (engl. *non-proprietary*),
- 3) format je dobro dokumentiran popratnom dokumentacijom (engl. *well documented*),
- 4) format je neovisan o platformi ili u potpunosti interoperabilan (engl. *platform independent*.) i
- 5) format je ili nekomprimiran ili je komprimiran bez gubitka podataka (engl. *uncompressed or compressed using a lossless technique*).

Istraživanje izričito navodi tih pet kriterija kao najvažnije karakteristike koje treba tražiti u formatu koji želimo koristiti u svrhu dugoročnog očuvanja i stvaranja preporuka ili politika oko odabira prikladnih formata za dugoročnu pohranu. Međutim, mnogi projekti i institucije koje provode projekte dugoročnog očuvanja navode kako je nekada teško pronaći format koji zadovoljava sve kriterije kod očuvanja različitih vrsta gradiva te je stoga potrebno naći prihvatljive alternative. Jedan od primjera je kriterij vlasništva nad formatom gdje se problem što neki format ima vlasnika može ublažiti činjenicom da ima javno objavljenu specifikaciju, da je postao industrijski standard ili da je prihvaćen od zajednice.

²¹ The InterPARES 2 Project. General Study 11 Final Report: Selecting Digital File Formats for Long-Term Preservation, 2007.

ZAKLJUČAK

Svaki projekt koji se bavi dugoročnim očuvanjem mora voditi računa o formatima za pohranu, a samim time i o kriterijima za odabir formata. Kod odabira formata za dugotrajno očuvanje elektroničkog gradiva u baštinskim ustanovama, potrebno je uzeti u obzir mnoga pitanja. Unatoč velikim naporima, još uvijek je problematično odabrati najprikladniji format za dugoročnu pohranu ili samo razlikovati prihvatljive od neprihvatljivih formata.

Prvi problem predstavlja sam pojam formata zapisa koji ima različite definicije u različitoj literaturi. U ovom radu pod pojmom formata zapisa smatramo prethodno ustanovljenu strukturu podataka digitalnih objekata koja ovisi o vrsti zapisa te je stvorena radi olakšanja pohrane, dohvata, obrade, predstavljanja i/ili prijenosa podataka pomoću softvera. Tom definicijom pokušalo se obuhvatiti sve aspekte pojma kroz definiranje razine strukture kao i praktične razine korištenja.

Uz pojam formata zapisa javlja se i povezano područje kategorizacije samih formata prema različitim kriterijima gdje prema kriteriju vlasništva možemo razlikovati zaštićene (engl. *proprietary file formats*), dostupne (engl. *available file formats*) i otvorene formate (engl. *open file formats*), prema kriteriju svrhe korištenja možemo razlikovati formate očuvanja (engl. *preservation formats*) i formate pristupa (engl. *access format*) dok se prema kriteriju prihvaćenosti u zajednici i podrške također pojavljuju pojmovi standardnog formata i stabilnog formata kojima se želi naglasiti otpornost formata na promjene tijekom vremena.

Središnji dio rada bavio se kriterijima za odabir formata za dugoročnu pohranu gdje se pokazalo kako postoji niz pristupa i mogućih kriterija koji se mogu koristiti.

U svjetlu formuliranja preporuka i smjernica ovdje donosimo prilagođene preporuke InterPARES 2 projekta²² koji je analizirao 20 digitalnih repozitorija te niz povezanih projekata dugoročnog očuvanja. Temeljene na iscrpnoj analizi, preporuke veoma dobro sumiraju aspekte

²² The InterPARES 2 Project. General Study 11 Final Report: Selecting Digital File Formats for Long-Term Preservation, 2007.

koji se mogu preporučiti svakoj instituciji koja kreće u proces razvijanja i implementiranja politika odabira formata za dugoročnu pohranu elektroničkog gradiva:

1. **Razjasniti terminologiju.** Jasno definirati što znače termini poput otvorenog formata, stabilnog formata, standardnog formata ili dobro dokumentiranog formata u samom dokumentu kojim se definira politika dugoročnog očuvanja.
2. **Razlikovati vrste formata zapisa.** Jasno razlikovati datotečne formate (engl. *file formats*), kontejner formate (engl. *container formats*) i *tagged* formate (npr. XML) i osigurati kako se njihove karakteristike poput verzije ili načina kodiranja u potpunosti razumiju i specificiraju.
3. **Izabrati odgovarajuće formate.** Odabrati formate koji su u širokoj primjeni, čije vlasništvo nad formatom nije zaštićeno i koji su neovisni o platformi.
4. **Definirati stupanj kompresije.** Odlučiti da li je prihvatljiva kompresija datoteka i ako jest, specificirati dozvoljeni stupanj kompresije. Kada god je moguće, izabrati kompresiju bez gubitka podataka koja slijedi prihvaćene međunarodne standarde.
5. **Konzultirati struku.** Ako nije moguće odabrati format koji zadovoljava kriterije navedene u preporuci pod brojem 3, izabrati formate koji se koriste u drugim repozitorijima digitalne građe te surađivati unutar struke kako bi se razvili odgovarajući planovi očuvanja korištenih formata.
6. **Surađivati sa stvarateljima.** Gdje god je moguće, usko surađivati sa stvarateljima gradiva kako bi se osiguralo korištenje programa koji podržavaju formate navedene u preporuci broj 3.

Uz navedene preporuke, u svjetlu dugoročnog očuvanja, također donosimo komplementarne preporuke izvještaja *Knowing the Need*²³ koji je od 2006. do 2011. proveo analizu 30 knjižničnih i 56 arhivskih zbirki u Ujedinjenom Kraljevstvu i Irskoj, kako bi ocijenio njihove strategije dugoročnog očuvanja.

²³ Peach, C, Foster, J. *Knowing the Need: Optimising Preservation for Library and Archive Collections*, British Library Preservation Advisory Centre, London. 2013. URL: https://www.bl.uk/aboutus/stratpolprog/collectioncare/publications/reports/knowning_the_need.pdf (18.7.2016.)

Izveštaj posebno izdvaja tri područja ključna za poboljšanje dugoročne uporabljivosti zbirke koji se mogu primijeniti kako na analogne tako i na digitalne zbirke:

1. **Ustanove moraju težiti što boljem razumijevanju svojih zbirki.** Poseban naglasak treba staviti na područja mogućih ranjivosti građe i gradiva, budućeg korištenja kao i postojećeg stanja i uporabljivosti, što je moguće postići kroz definiranje razina opisa, jasnog definiranja rokova čuvanja, provođenje procjene stanja zbirke i građe te strateškog planiranja buduće uporabe građe.
2. **Uočene slabosti i nedostaci u praksi očuvanja, posebice u području pohrane, uvjeta čuvanja te planova u slučaju katastrofe trebaju biti otklonjene.** Prostorije za pohranu trebaju biti redovito održavane, građa odgovarajuće pohranjena s obzirom na različite formate, a sigurnosni planovi redovito razmatrani te osoblje upoznato s njihovom implementacijom. Također, ustanove trebaju redovito pratiti nove smjernice koje im mogu pomoći u poboljšanju energetske učinkovitosti i zaštiti najugroženije građe.
3. **Surogati trebaju biti razmotreni kao strategija očuvanja.** Ustanova mora jasno definirati uvjete pod kojima digitalni objekti mogu biti surogat za dugoročno očuvanje analogne građe te dobro razumjeti potencijale programa digitalizacije u poboljšanju pristupa i korištenja zbirke.

Ovaj rad pruža okvir za kreiranje odgovarajuće strategije i politike pri odabiru formata datoteka za dugotrajno očuvanje i pristup digitalnim sadržajima. Pregled kriterija opisanih u ovom radu može koristiti arhivima, knjižnicama, muzejima i ostalim baštinskim ustanovama kada se susreću s pitanjima razvijanja sustava za dugoročnu pohranu i očuvanje građe. Praktična pitanja na razini implementacije i odabira formata odluka je koju svaka baštinska institucija mora donijeti u svom lokalnom kontekstu, uzimajući u obzir potrebe svojih korisnika, dostupne kadrove i financijske resurse.²⁴ Pojavom novih formata neminovno je i mijenjanje kriterija njihova odabira kako bi bili odabrani formati koji najbolje služe svojoj svrsi, ali stalna svijest o važnosti praćenja uloge formata zapisa sa svim tehničkim, arhivskim, proceduralnim i

²⁴ Park, E. G. & Oh, S. Examining Attributes of Open Standard File Formats for Long-term Preservation and Open Access. *Information Technology and Libraries*, 31, 4(2012), str. 44-65.

financijskim pitanjima koje to uključuje važan je element ciklusa upravljanja dugoročnim očuvanjem elektroničkog gradiva.

Criteria for selection of file formats for long-term preservation

Each institution dealing with long-term preservation must take into account the appropriate file formats for storage and preservation of digital materials. This paper provides a framework for creating appropriate strategies and policies in selecting file formats for long-term preservation and access to digital content. The criteria described in this paper can be used in archives, libraries, museums and other heritage institutions when faced with the issues of developing a system for long-term storage and preservation of digital materials.

The first part of the paper analyses the very concept of formats and related areas of format categorization. The central part of the paper analyzed a number of possible approaches and criteria that can be used for selection of formats for long-term storage of digital materials. The final part of the paper formulated recommendations and guidelines based on recommendations of InterPARES 2 project that summarize the aspects of format selection that any institution participating in the process of developing and implementing policies for long-term preservation should take into consideration: clarify terminology; distinguish between file formats; choose appropriate formats; define the degree of file compression; consult the professionals; cooperate with the creators.

The criteria may change over time, as is necessary for any format to adequately serve its purpose. Maintaining its quality may be an ongoing task that cultural heritage institutions should take into account at all times and each institution should consider implementing a file format throughout the entire life cycle of digital resources, from managerial, technical, procedural, archival, to financial issues for the purpose of long-term preservation and persistent access.

Keywords: file format, long-term preservation, selection criteria, heritage institutions, digital materials

Popis literature

Barnes, I. Preservation of Word Processing Documents. *Australian Partnership for Sustainable Repositories (APSR)*. 2006.

Barve, S., File Formats in Digital Preservation. *International Conference on Digital Libraries (ICSD)* (Bangalore) 2007. str. 239-248.

Becker, C., Rauber, A. Decision Criteria in Digital Preservation : What to Measure and How. *Journal of the American Society for Information Science and Technology (JASIST)*, 6, 62, 6(2011). str. 1009-1028.

Hrvatski državni arhiv. *Arhivska služba*. 2013.

URL: <http://www.arhiv.hr/arhiv2/Arhivskasluzba/index.htm> (18.7.2016.)

Mihaljević, M., Mihaljević, M., Stančić, H. *Arhivistički rječnik*. Zagreb: Zavod za informacijske studije. 2015.

Park, E. G., Oh, S. Examining Attributes of Open Standard File Formats for Long-term Preservation and Open Access. *Information Technology and Libraries*, 31, 4(2012), str. 44-65.

Peach, C, Foster, J. Knowing the Need: Optimising Preservation for Library and Archive Collections. *British Library Preservation Advisory Centre*, London. 2013. URL: <http://www.bl.uk/blpac/pdf/ktn.pdf> (18.7.2016.)

PREMIS Working group. Data Dictionary For Preservation Metadata. *OCLC Online Computer Library Center, Inc. i RLG*. 2005.

Rauch, C., Krottmaier, H., Tochtermann, K. File-Formats for Preservation: Evaluating the Long-Term Stability of File Formats. *Proceedings ELPUB2007 Conference on Electronic Publishing* (Beč). 2007.

Rog, J., Van Wijk, C. Evaluating File Formats for Long-Term Preservation. *The Hague: National Library of the Netherlands*. 2008. str. 83-90.

Sahu, D. K. Long Term Preservation: Which File Format to Use. *Workshops on Open Access & Institutional Repository* (Chennai). 2004.

Sullivan, S. J. An Archival/records Management Perspective on PDF/A. *Records Management Journal*, 16, 1(2006). str. 51-56.

This is a preprint version of a published paper. For citing purposes please use:

Ivanjko, Tomislav; Lončarić, Ksenija. Kriteriji za odabir formata za dugoročnu pohranu elektroničkoga gradiva // Arhivi i domovinski rat / Babić, Silvija (ur.). Zagreb : Hrvatsko arhivističko društvo, 2016. 413-424

The InterPARES 2 Project. *General Study 11 Final Report: Selecting Digital File Formats for Long-Term Preservation*, 2007.

The InterPARES 2 Project. *The InterPARES 2 Project Dictionary*. 2016.