

SVEUČILIŠTE U ZAGREBU
FILOZOFSKI FAKULTET
ODSJEK ZA INFORMACIJSKE I KOMUNIKACIJSKE ZNANOSTI

MARKO ANDREIS

PREPORUČENI ARHIVSKI FORMATI ZAPISA DIGITALNE GRAĐE S OBZIROM
NA NJEZINU VRSTU I NAMJENU

DIPLOMSKI RAD

Mentor:

dr. sc. Hrvoje Stančić, red. prof.

Zagreb, rujan 2018.

*Zahvaljujem svojoj obitelji
na strpljenju i podršci tijekom
protekle tri godine studiranja*

*Zahvaljujem mentoru Hrvoju Stančiću
na savjetima i pomoći prilikom
izrade ovog diplomskog rada*

Sadržaj

Sažetak	1
Abstract	1
1 Uvod.....	2
2 Općenito o digitalizaciji	4
3 Formati zapisa digitalne građe	10
3.1 Formati za tekstualno gradivo	12
3.1.1 BITS format.....	15
3.1.2 EPUB format	15
3.1.3 PDF formati	16
3.1.4 NISO/JATS format	19
3.1.5 MusicXML format.....	19
3.1.6 MEI format	20
3.2 Formati za nepokretne slike.....	29
3.2.1 TIFF format	31
3.2.2 JPEG2000 format	32
3.2.3 PNG format.....	32
3.2.4 JPEG/JFIF format	33
3.2.5 DNG format.....	33
3.2.6 BMP format	33
3.2.7 GIF format	34
3.2.8 SVG format.....	34
3.2.9 DXF format.....	34
3.2.10 EPS format.....	35
3.2.11 SHP format	36

3.2.12	GeoTIFF format.....	36
3.2.13	GeoPDF format.....	37
3.2.14	GeoJPEG2000 format.....	37
3.3	Formati za zvukovnu građu	44
3.3.1	WAVE format.....	46
3.3.2	Broadcast WAVE format.....	47
3.3.3	DSD format.....	47
3.3.4	MP3 format.....	47
4	Zaključak.....	51
5	Popis tablica	56
6	Literatura.....	57

Sažetak

Cilj ovog diplomskog rada je prikazati najčešće vrste formata digitalnih zapisa za tri najčešće vrste gradiva koje bi se mogle naći u projektima digitalizacije knjižnica i arhiva. To su tekstualno gradivo, kao najzastupljenije, zatim nepokretne slike i zvukovno gradivo.

Kao temelj za analizu poslužile su smjernice za digitalizaciju Kongresne knjižnice i njihove najnovije preporuke o preporučenim formatima (LOC Recommended Formats Statement 2017-2018).

Neovisno o vrsti gradiva koje će se digitalizirati u svrhu dugoročnog očuvanja, nekoliko kriterija je zajedničko prilikom odabira najpovoljnijeg digitalnog formata: rasprostranjenost i usvojenost formata, njegova dokumentiranost, neovisnost o licencama i drugim intelektualnim pravima te količina metapodataka koja opisuje sadržaj datoteke.

Uz ove kriterije svaka vrsta gradiva postavlja i dodatne kriterije specifične za tu vrstu gradiva.

Ključne riječi: digitalni formati, dugoročno očuvanje.

Abstract

The goal of this thesis is to show most common types of digital file formats for three most common types of content, likely to be subject of digitization projects in libraries and archives. These are text as the most likely one, still pictures and sound.

The analysis was based on Library of Congress guidelines for digitization and LOC Recommended Formats Statement 2017-2018 document.

Independently of the content type to be digitized for long term preservation they all share some common criteria that must be met when deciding on the most appropriate format – the format has to be widespread and adopted, documentation must exist and has to be maintained, the format has to be free to use in terms of intellectual property and licenses and the amount of metadata that describe content details has to be sufficient for the purpose.

In addition to these criteria, every content type sets additional set of criteria that is specific for that content type.

Key words: digital formats, long term preservation.

1. Uvod

Od početka ljudske pismenosti čovjek je stvarao zapise kojima je cilj bio zabilježiti, i za buduća pokoljenja sačuvati znanja i vjerovanja tadašnjeg vremena, kao i prenijeti znanja i vjerovanja predaka. Iako ljudi s početka razvoja pisane riječi vjerojatno nisu razmišljali o problemima dugoročnog očuvanja vlastitih zapisa na način na koji mi to danas činimo, zapravo su se suočavali s istim problemima: kako medij učiniti što trajnijim i kako informaciju učiniti razumljivom i nakon proteka brojnih godina. I dok su se neki najraniji mediji pokazali dovoljno trajnima te su dobro očuvani i čitljivi danas, razumijevanje njihovih zapisa modernom je čovjeku predstavljalo velik izazov, budući da se znanje o tumačenju simbola koji su na te medije zapisani izgubilo tijekom povijesti.

U današnje vrijeme brz razvoj tehnologije za posljedicu je imao stvaranje globalne informacijske infrastrukture i globalizaciju informacijskih sustava. Ovi sustavi omogućili su globalnu komunikaciju putem računala i stvaranje i razmjenjivanje golemih količina informacija.¹ Kako bi se mogle razmjenjivati putem informacijske infrastrukture, ove informacije moraju postojati u obliku prilagođenom toj informacijskoj infrastrukturi, tj. moraju biti u digitalnom obliku. Informacije, odnosno objekti u kojima su sadržane mogu biti izvorno digitalni (engl. *born-digital*, *natively digital*, *digital-first*, *digital-exclusive*) ili mogu nastati iz analognih objekata postupkom digitalizacije.

S gledišta dugoročnog očuvanja digitalni objekti suočeni su sa sličnim problemima s kojima se susreću objekti u svojem nedigitalnom obliku (pisani dokumenti, gramofonske ploče, video i magnetofonske vrpce, trodimenzionalni objekti i sl.) – trajnost medija, razumljivost zapisa i način čitanja zapisa. Sva ova tri problema obilježavaju specifičnosti vezane uz određeni moderni medij.

Mediji koji su se koristili prije pojave modernih medija namijenjenih strojnom čitanju kroz povijest su pokazali svoje prednosti i nedostatke u odnosu na trajnost. Tako su se kamen, glinene pločice, pergament i papir pokazali kao trajni mediji koji su se uspjeli sačuvati do danas, dok su zapisi na platnu manje trajni. O trajnosti modernih medija možemo samo teorijski raspravljati na temelju simuliranog starenja, budući da je prošlo premalo vremena od njihove pojave do danas.

¹ Stančić, H., Digitalizacija. Zagreb : Zavod za informacijske studije, 2009.

Ranije je spomenuto da je razumijevanje prvih ljudskih zapisa današnjem čovjeku predstavljalo izazov. S istim izazovom susrest će se nalaznik nekog danas modernog medija u budućnosti budući da je način zapisivanja zapisa na medij uvelike ovisan o samom mediju. Naravno, uz pretpostavku da taj nalaznik iz budućnosti taj medij bude mogao na ovaj ili onaj način pročitati. Ukoliko se ne sačuva dokumentacija o samom načinu zapisivanja, do otkrivanja značenja zapisa mogu proći desetljeća.

Za razliku od ranijih medija kod kojih je način čitanja bio isključivo ljudsko čitanje, mnogi moderni mediji namijenjeni su strojnom čitanju. Želimo li u budućnosti s lakoćom pročitati medij sa zapisom stvorenim u današnje vrijeme, morali bismo raspolagati tehnologijom koja takav medij može pročitati. Uz rapidan razvoj tehnologije, mediji vrlo brzo nakon svoje pojave postaju zastarjeli, te je nemoguće sa sigurnošću tvrditi da će danas upotrijebljeni medij biti moguće pročitati za dvadeset ili čak manje godina, bez obzira na to koliko dobro sâm medij bio očuvan.

Spomenuti rapidni razvoj tehnologije i sve mogućnosti koje on pruža dovodi u drugi plan pitanje dugoročnog očuvanja digitalnog gradiva, tako da neki znanstvenici čak najavljuju dolazak digitalnog srednjeg vijeka, tj. društva bez ostavštine (Kuny, 1997.; Ilshammar, 2001.).²

Glavna tema ovog diplomskog rada je problem razumljivosti zapisa, odnosno formata zapisa i to u zavisnosti o sadržaju zapisa.

U radu će se usporediti različiti formati zapisa namijenjeni istoj vrsti sadržaja u svjetlu njihove prikladnosti za dugotrajno očuvanje tj. arhiviranje.

Na kraju uvoda potrebno je razjasniti jedno terminološko pitanje. Naime, kroz cijeli diplomski rad provlače se dva pojma – gradivo i građa. Strogo gledano, gradivo se odnosi na izvorni i reproducirani dokumentirani materijal koji se čuva u arhivima, a građa se odnosi na sve ono što knjižnice kroz svoje djelovanje skupljaju, obrađuju i korisnicima stavljaju na raspolaganje. Međutim zbog toga što je tema ovog rada podjednako povezana s arhivistikom i bibliotekarstvom, ova dva pojma koristit će se kao istoznačni.

² Runnardotter, M. On organizing for digital preservation : Doctoral thesis. Luelå : Luelå University of Technology, 2009.

2. Općenito o digitalizaciji³

Pojam digitalnoga danas susrećemo u svim aspektima života, od slobodnog vremena i zabave, preko svakodnevnog života građana, ponude kulturnih ustanova pa do medicine i poslovanja velikih korporacija i državnih institucija. U svakom od ovih slučajeva predmet obrade je općenito neki digitalni zapis. To može biti poruka elektroničke pošte ili poruka na nekoj od brojnih platformi za razmjenu poruka (Viber, Whatsapp, Skype, SMS, MMS...), mrežna stranica, elektronička knjiga, financijska transakcija, upravni i administrativni akti (izvodi iz matičnih knjiga, uvjerenja, vlasnički listovi, izvaci iz katastra...), medicinska dokumentacija (nalazi, RTG, CT, MR snimke...), poslovna dokumentacija (ulazni ili izlazni računi, otpremnice, dostavnice, dokumenti o kontroli kvalitete i otpuštanju pojedinih serija proizvoda na tržište...) ili umjetnička djela predstavljena javnosti u virtualnom prostoru, kao i arhivski materijali.

Zapis je dokument ili objekt (uključujući bilo koji elektronički oblik) koji jest ili je bio sačuvan zbog bilo koje informacije ili sadržaja koje posjeduje, ili značenja koje može biti iz njega izvedeno, ili zbog njegove veze s nekim događajem, osobom, okolnošću ili stvari.⁴ Iz ove definicije vidljivo je da zapisi, osim što mogu biti fizički, mogu biti i elektronički, tj. digitalni. Elektronički zapisi prema svom nastanku mogu biti izvorno digitalni ili mogu nastati iz analognih objekata postupkom digitalizacije.

Bez obzira na to o kakvom se zapisu radi, svaki zapis ima svoji životni vijek kroz koji ga se mora održavati kako bi u trenutku kada ga se treba pročitati on bio dostupan i vjerodostojan. Ovo vrijedi jednako za fizičke zapise, dakle one na medijima koji su čitljivi bez upotrebe računala, kao i za digitalne zapise, bez obzira na način njihovog nastanka.

Ukoliko elektronički zapis nije izvorno digitalan govorimo o digitalizaciji kao postupku pretvorbe materijalnih zapisa u digitalni oblik. Razlozi zbog kojih se upuštamo u ovaj postupak mogu biti različiti, a najčešći su zaštita izvornika, povećanje dostupnosti i mogućnosti korištenja građe, stvaranje nove ponude, odnosno usluga korisnicima ili upotpunjavanje postojećeg fonda, a u novije vrijeme to je i digitalizacija na zahtjev. Svaki od ovih razloga ima svoje specifičnosti o kojima treba voditi računa prilikom planiranja i izvođenja projekta digitalizacije.

³ Stančić, Digitalizacija, n. dj.

⁴ The Archives Act, Australia, 1983., URL: http://www7.austlii.edu.au/cgi-bin/viewdoc/au/legis/cth/consol_act/aa198398/s3.html. (12. ožujka 2018.)

Digitalizacija u svrhu zaštite izvornika provodi se zbog dva osnovna aspekta zaštite. Kroz prvi se korisniku nudi elektronička inačica zapisa umjesto klasične (papirne ili neke druge), a kroz drugi digitalizirana inačica može poslužiti kao sigurnosna kopija u slučaju oštećenja ili uništenja izvornika.

Digitalizacija u svrhu povećanja dostupnosti gradiva jasna je sama po sebi. Jedan primjerak izvornika može biti ponuđen samo jednom korisniku, dok njegova digitalizirana inačica može biti dostupna istovremeno velikom broju korisnika. Kroz povećanje dostupnosti gradiva institucije mogu povećati broj svojih korisnika, a u slučaju digitalizacije ključnog gradiva važnog za kulturnu baštinu ujedno doprinosi promicanju države ili naroda na globalnoj razini.

Stvaranje nove ponude jedan je od glavnih razloga digitalizacije gradiva. Digitalizirani zapis popraćen pažljivo odabranim i točnim metapodacima omogućuje provođenje različitih analiza sadržaja, povezivanje sadržaja različitih izvora u virtualne zbirke, brzo pretraživanje punog teksta i razmjenu samih metapodataka.

Digitalizacija koja za cilj ima upotpunjavanje fonda provodi se nad gradivom koje nije u vlasništvu institucije, a digitalizirana inačica takvog gradiva može upotpuniti nepotpuni ili uništeni dio fonda institucije. Na taj način digitalizirani oblik ostaje jedini oblik u vlasništvu institucije.

I na kraju digitalizaciju na zahtjev valja promatrati kao jednu od dodatnih usluga koje institucija pruža te je kao takva slična digitalizaciji s ciljem stvaranja nove ponude. Statistike o tome koja građa je najčešće zatražena kroz ovaj oblik digitalizacije mogu poslužiti pri određivanju prioriteta digitalizacije prilikom pripreme projekata digitalizacije.

Planira li institucija digitalizirati svoju građu ili jedan njezin dio mora osmisliti projekt digitalizacije. Kroz projekt digitalizacije procjenjuju se svi aspekti postupka: materijalno-tehnički, financijski, sigurnosni, pravni i administrativni. Također, institucija mora biti svjesna da je sam postupak digitalizacije težak i često dugotrajan, te da završetkom digitalizacije ne završava posao oko održavanja digitaliziranih podataka.

Stančić tumači da se projekti digitalizacije u načelu provode kroz sedam faza pri čemu faze ne moraju teći slijedno.⁵

Prva faza projekta odnosi se na odabir gradiva koje će se digitalizirati. Na početku projekta digitalizacije institucija treba donijeti odluku o tome koje će gradivo

⁵ Stančić, Digitalizacija, n. dj.

digitalizirati. Taj odabir ovisi o vrsti ustanove, njezinom programu i ciljevima digitalizacije. Proces odabira gradiva provodi se kroz matricu odlučivanja. Ona donosi korake i njihov redoslijed te pitanja koja prilikom odabira gradiva treba razmotriti.

Odabir gradiva odvija se kroz četiri faze. Osim već spomenute faze donošenja općih kriterija odabira odabir provodimo kroz predlaganje gradiva, procjenu gradiva i određivanje prioriteta digitalizacije.

U procesu predlaganja gradiva članovi povjerenstva za odabir iznose mišljenje o tome treba li određene segmente gradiva uzeti u razmatranje za digitalizaciju i zašto. Rezultat predlaganja gradiva je popis gradiva predloženog za digitalizaciju i popis onog gradiva koje nije predloženo za digitalizaciju.

Na temelju predloženog gradiva za digitalizaciju potrebno je procijeniti koje gradivo digitalizirati. Popisi gradiva nastali postupkom predlaganja gradiva uspoređuju se i povjerenstvo donosi zajedničku odluku o tome koje će se gradivo digitalizirati.

Posljednji korak u fazi odabira gradiva za digitalizaciju je određivanje prioriteta digitalizacije. Budući da sâm postupak digitalizacije traje dugo prioritete treba odrediti vodeći računa o trima aspektima: vrijednosti, riziku i predviđenom korištenju gradiva.⁶

Nakon što je gradivo za digitalizaciju odabrano pristupa se drugoj fazi projekta – digitalizaciji. Digitalizirati se može tekstualno, slikovno, zvučno, video i trodimenzionalno gradivo. Vrsta gradiva i njegove fizičke dimenzije diktiraju odabir opreme i postupke njegove digitalizacije.

U trećoj fazi postupka digitalizacije, potrebno je provjeriti zadovoljava li kvaliteta digitaliziranog gradiva postavljene standarde, a uz to je najčešće potrebno provesti i dodatnu obradu. Dodatna obrada specifična je za svaku vrstu gradiva, a ono što je svima zajedničko je potreba dodavanja metapodataka koji olakšavaju pretraživanje datoteka.

Četvrta faza projekta opisuje zaštitu digitalnog gradiva nakon njegove obrade i potvrde njegove kvalitete. Budući da različiti projekti digitalizacije za ciljanu publiku imaju različite skupine korisnika potrebno je povesti računa o zaštiti gradiva, tj. o tome da gradivu mogu pristupiti samo oni koji za to imaju ovlaštenje i da ga korištenjem ne mogu mijenjati ili obrisati.

⁶ Vogt-O'Connor, D. Reformatting for Preservation and Access: Prioritizing Materials for Duplication // Conserve O Gram. 19/10(1995), <https://www.nps.gov/museum/publications/conserveogram/19-10.pdf>, 3. travnja. 2018.

Za zaštitu digitalnog gradiva postoje različiti mehanizmi, a zaštita se najčešće osigurava primjenom više njih. Najniža i najopćenitija razina zaštite je ona na razini operativnih sustava i mreže. Sljedeća kategorija mehanizama zaštite djeluje na razini datotečnog sustava na kojoj se određuje kojim datotekama pojedini korisnici smiju pristupiti bez da povrede prava vlasnika i distributera gradiva, a kojima ne. Mehanizam kriptiranja (šifriranja, zakrivanja) je sljedeća viša razina zaštite. Ovim mehanizmom gradivo se čini nečitljivim svima osim onim korisnicima koji posjeduju ispravan ključ za dekriptiranje (dešifriranje, raskrivanje). Proširenje mehanizma kriptiranja je mehanizam postojanog kriptiranja kojim se dekriptira samo onaj dio gradiva koji se trenutno koristi, dok ostali dijelovi ostaju kriptirani. Na kraju, mehanizmom digitalnih potpisa i digitalnih vodenih žigova u gradivo se ugrađuju informacije o vlasništvu.

Pohrana i prijenos digitalnog gradiva peta je faza projekta digitalizacije. Uz digitaliziranu građu, danas digitalno gradivo nastaje kao rezultat svih poslovnih procesa bilo koje institucije. Stoga je i taj dio digitalnog gradiva potrebno uzeti u obzir kada se procjenjuje potreba za kapacitetom prostora za pohranu i samim sustavom za pohranu. Osnovni kriteriji za odabir sustava za pohranu su dugovječnost medija (podatak koji govori koliko će se odabrana vrsta medija zadržati na tržištu), trajnost medija (koliko je rok trajanja svakog pojedinog medija), visok kapacitet, mala cijena, široka prihvaćenost i izravnost, odnosno poluizravnost sustava.

Prilikom planiranja projekta digitalizacije u šestoj fazi predviđa se načine pregledavanja i korištenja gradiva. Nije svejedno hoće li se gradivo pregledavati isključivo na zaslonu računala ili će ga se ispisivati na pisačima. Zatim je važan čimbenik gdje će se gradivo pregledavati – hoće li to biti samo unutar mreže institucije, dok su putem Interneta dostupni samo metapodaci, ili će cjelokupno gradivo biti dostupno na Internetu. Ukoliko institucija planira omogućiti ispis gradiva na zahtjev morat će osigurati kvalitetne uređaje za ispis. I dok je to za tekstualno gradivo relativno jednostavno i jeftino, ispis slikovnog gradiva zahtjeva veća ulaganja. I na kraju, valja povesti računa i o tome da krajnji korisnik nije uvijek čovjek, već to mogu biti i različiti programi i programski agenti, pa je potrebno osigurati odgovarajuće sučelje putem kojeg će se gradivo izložiti takvim „korisnicima“.

Sedma faza projekta digitalizacije je kontinuirani proces održavanja gradiva. Digitalno gradivo potrebno je održavati zbog zastarijevanja elektroničkih sustava, medija i zapisa.

Gradivo u klasičnom obliku bit će dostupno onoliko koliko i medij od kojega je napravljeno. Za knjigu to može biti i nekoliko stotina godina ukoliko je ona pohranjena u uvjetima koji na nju ne djeluju štetno. Međutim pitanje je hoćemo li moći pročitati digitalni zapis načinjen prije dvadesetak godina. Prepreka tome može biti već i činjenica da više ne postoje ispravni čitači medija na kojima se zapis nalazi. Ako se takav čitač i uspije nabaviti, može se dogoditi da današnja računala nemaju odgovarajuće fizičko sučelje za spajanje tog čitača. Ako se i ovaj problem savlada, računalu je potreban pokretački program, tzv. *driver* za upravljanje čitačem, koji za zastarjelu opremu možda nije više dostupan. Zatim se može dogoditi da na računalu ne postoji program koji bi uspješno interpretirao sadržaj zapisa, uz pretpostavku da je sam medij dobro očuvan i da zapis na njemu nije promijenjen od trenutka kad je zapisan.

Općenito, problemi koji se javljaju prilikom održavanja digitalnoga gradiva mogu se grupirati u pet skupina.⁷ **Problem pregleda gradiva** vidljiv je kada se pokušava pregledati neko starije elektroničko gradivo, a programi novije generacije ne mogu ispravno protumačiti taj zapis i format u kojem je zapisan.

Problem kodiranja gradiva javlja se prilikom sažimanja gradiva. Prilikom sažimanja veličina datoteke se može značajno smanjiti, što omogućava veću brzinu prijenosa gradiva i omogućuje pohranu veće količine gradiva na mediju. Međutim sažimanje iz gradiva uklanja neke elemente koje ljudsko oko ne razaznaje i danas ne možemo znati hoće li nam uklonjeni element u budućnosti biti ključan za analizu tog gradiva.

Problem međusobne povezanosti sadržaja javlja se zbog činjenice da se građa koju korisnik pregledava nerijetko sastoji od više različitih datoteka koje su međusobno povezane tako da korisniku izgledaju kao jedinstvena cjelina. Na primjer, slike unutar digitalnoga arhiva mogu se nalaziti u jednoj datoteci, a naslov i metapodaci u drugoj. U slučaju mrežnih stranica često se nailazi na poveznice prema drugim stranicama te se javlja problem prepoznavanja svih poveznica i spremanja svih stranica na koje te poveznice upućuju. A budući da i te stranice sadrže svoje poveznice postavlja se pitanje koliko duboko treba ići s pohranom sadržaja poveznica. Zatim, elektroničke poruke koje također mogu biti dio građe često se sastoje od tijela poruke i privitaka – slikovnih, zvučnih, tekstualnih i drugih datoteka. U svim ovim slučajevima potrebno je sastavne dijelove građe organizirati tako da se zadrže njihove međusobne veze.

⁷ Stančić, Digitalizacija, n. dj.

Prilikom arhiviranja gradiva, bez obzira radi li se o digitaliziranom ili digitalno stvorenom gradivu, javlja se **problem ovlasti za arhiviranje**. I dok su kod klasičnog gradiva uglavnom jasno određene odgovornosti za arhiviranje, one za gradivo nastalo u digitalnom obliku još uglavnom nisu definirane. U svakom slučaju, ovaj problem daleko je kompleksniji u slučaju digitalnog gradiva u odnosu na klasično gradivo zbog stalnih tehnoloških inovacija te novih oblika mrežne komunikacije.

Budući da se programi za interpretaciju digitalnih zapisa stalno mijenjaju događa se da novija inačica programa više ne prepoznaje format zapisa nekog starijeg dokumenta, unatoč tome što je njegov fizički zapis ostao nepromijenjen. Kako bi se to spriječilo potrebno je stariji zapis konvertirati u novi format. Prilikom **konverzije zapisa** može doći do promjena u samom izgledu gradiva. Zbog toga je bitno da se uz gradivo koje nastaje u elektroničkom obliku zapisuju metapodaci koji opisuju podatke o njegovoj strukturi i programskom okruženju u kojem je nastalo kako bi se ono moglo lakše rekonstruirati.

Ukoliko se želi rizike od pojave spomenutih problema svesti na najmanju moguću mjeru potrebno je neprestano provoditi proces održavanja gradiva. U načelu ovo održavanje može se svesti na tri postupka od kojih je svaki usmjeren na jedan od problema. Ti procesi su osvježavanje medija, migracija zapisa i emulacija programske okoline.

Osvježavanje medija je postupak u kojem se ne mijenja fizički zapis elemenata građe, već se građa sa starih, dotrajalih medija periodički kopira na nove medije iste generacije. Ukoliko zapis kopiramo na medij novije generacije govorimo o **migraciji zapisa**.

Za razliku od osvježavanja medija, odnosno migracije zapisa, **konverzija zapisa** mijenja fizički zapis elemenata građe. Ovim postupkom se građa zapisana u zastarjelom formatu zapisa konvertira u noviji format čitljiv programima novije generacije.

Emuliranjem programske okoline zapravo na nekoj tehnološki naprednijoj računalnoj platformi simuliramo elektroničko okruženje neke zastarjele platforme, tako da se u takvom simuliranom okruženju može pokrenuti neki program koji moderne platforme više ne mogu pokretati. Kako bi se emulator mogao uspješno napraviti potrebno je uz izvorni dokument pohraniti i podatke o aplikacijskoj okolini i operativnom sustavu na kojem je on nastao, specificirati parametre emulatora hardvera te metapodatke o samoj građi.

3. Formati zapisa digitalne građe

Tema ovog diplomskog rada najuže je povezana s konverzijom zapisa, odnosno odabirom najpovoljnijeg formata datoteke za pohranu digitalnoga gradiva.

Danas je dostupan niz različitih formata za pohranu digitalnoga sadržaja za različite vrste gradiva. Međutim, nisu svi formati prikladni za dugoročno očuvanje građe.

Rad će predstaviti pregled formata za tekstualno, slikovno i zvukovno gradivo te usporediti preporučeni format s drugim dostupnim formatima i objasniti prednosti preporučenog formata u odnosu na druge.

Kada govorimo o dugoročnom očuvanju građe i formatima namijenjenim za tu svrhu potrebno je navesti definicije dugoročnog očuvanja i formata.

Dugoročno očuvanje definira se kao skup aktivnosti potreban za održavanje građe u upotrebljivom obliku kroz dugi period. Iako izraz „dugi period“ nije egzaktno definiran, smatra se da je on dovoljno dug da za njegova trajanja na građu mogu utjecati tehnološke i društvene promjene. Zbog toga je za pohranu sadržaja koji želimo dugoročno očuvati potrebno odabrati format koji će se s vremenom mijenjati kontrolirano i za koji se očekuje da će u svojim osnovnim tehničkim elementima biti prisutan nakon duljeg perioda. To ne znači da se neće tijekom vremena morati provoditi postupak konverzije podataka iz jedne inačice istog formata u neku noviju, ali pravilnim odabirom barem nastojimo osigurati da taj postupak bude što je moguće lakši.

Format se definira kao skup informacijskih paketa koji se mogu pohraniti kao podatkovne datoteke ili slati putem mreže kao protočne podatke (engl. *data stream*).⁸

Neki od dostupnih formata svojstveni su samo nekim računalnim platformama, samo nekom hardveru ili samo nekim aplikacijama i ta specijaliziranost čini ih neprikladnima za dugoročno očuvanje. Drugi formati, koliko god izgledali prikladno, nisu široko usvojeni ili su podložni intelektualnim pravima i licencama. Treći su možda slobodni za korištenje, ali su tek razvijeni i ne zna se koliko će postojati i kontroliran biti njihov razvoj. Najčešće se radi o nekoliko čimbenika koji određeni format eliminiraju iz izbora za najpogodniji format.

Općenito, pred format zapisa namijenjen dugoročnom očuvanju gradiva postavlja se sedam tehničkih zahtjeva.

⁸ Arms C. R.; Fleischhauer C.; Murray, K. Sustainability of Digital Formats: Planning for the Library of Congress Collections. URL: https://www.loc.gov/preservation/digital/formats/intro/format_eval_rel.shtml. (24. svibnja 2018.)

Otkrivenost je stupanj do kojega postoji potpuna specifikacija i alati za tehničku validaciju integriteta te njihova dostupnost onima koji stvaraju i održavaju digitalne sadržaje.

Usvojenost opisuje u kojoj mjeri format već koriste primarni tvorci, raspačavatelji ili korisnici informacijskih resursa. Oni uključuju upotrebu formata kao glavnog formata, kao formata za isporuku sadržaja i kao formata za međusobnu razmjenu između informacijskih sustava.

Transparentnost opisuje koliko je digitalni zapis otvoren izravnoj analizi pomoću najosnovnijih alata, npr. čitanje putem programa za uređivanje isključivo teksta.

Samodokumentiranje obuhvaća osnovne opisne, tehničke i druge administrativne metapodatke.

Vanjske ovisnosti govore o tome koliko određeni format ovisi o hardveru, operativnom sustavu ili aplikacijama za prikaz ili korištenje, te o predviđenoj složenosti upravljanja ovim ovisnostima u budućim tehničkim okruženjima.

Utjecaj патената kazuje u kojoj mjeri će se mogućnost institucije da gradivo održava u određenom formatu ograničavati patentnim pravima.

Postojanje **mehanizama tehničke zaštite**, poput šifriranja, moglo bi spriječiti očuvanje sadržaja.

Osim tehničkih zahtjeva, format se procjenjuje i kroz aspekt kvalitete i funkcionalnosti. Kvaliteta i funkcionalnost predstavlja sposobnost formata da prikaže značajna obilježja građe potrebna sadašnjim i budućim korisnicima. Zahtjevi koji se postavljaju u okviru kvalitete i funkcionalnosti variraju ovisno o građi kojoj je format namijenjen.

Digitalnu građu, a time i formate za njenu pohranu može se prema vrsti podijeliti u sljedeće skupine: tekstualne, zvuk, nepokretne slike i pokretne slike.

Formate za pohranu i dugoročno očuvanje digitalne građe promotrit će se na primjeru američke Kongresne knjižnice. Kongresna knjižnica donijela je 2014. godine Izjavu o preporučenim formatima i svake godine ova Izjava se revidira i nadopunjuje.⁹

Brojni formati za pohranu dokumenata koriste se XML (engl. *eXtensible Markup Language*) jezikom pa će stoga on u nastavku biti ukratko opisan.

⁹ Library of Congress Recommended Formats Statement : 2017-2018. Washington DC : Library of Congress. Dostupno na: <https://www.loc.gov/preservation/resources/rfs/RFS%202017-2018.pdf>. (24. svibnja 2018.)

XML jezik je označni jezik osmišljen za kodiranje i razmjenu dokumenata u formatu koji je čitljiv i ljudima i strojevima. Funkcija XML jezika u samom formatu za pohranu dokumenata može biti višestruka, od metapodataka do cijelog sadržaja dokumenta. Ovaj jezik razvija *World Wide Web Consortium* (W3C) i u potpunosti je besplatan i javan.

Osnovni elementi XML jezika su oznake unutar kojih se nalazi sadržaj. XML jezik razlikuje tri osnovne vrste oznaka: početne, završne i oznake bez sadržaja. Početna i završna oznaka dolaze u paru, a između njih se nalazi sadržaj. Oznake bez sadržaja mogu stajati samostalno. Oznake najčešće počinju znakom „<“, a završavaju znakom „>“.

XML element je zapravo sadržaj koji se smješta između početne i završne oznake. Elementi mogu sadržavati i druge ugniježdene elemente.

Atributi su parovi naziva i njihovih vrijednosti i stavljaju se unutar početne oznake ili unutar oznake bez sadržaja. Oni opisuju dodatne značajke elementa poput npr. veličine slova, boje, reference i sl.

Pravila upotrebe oznaka i atributa opisana su tzv. shemom ili gramatikom. Najstariji oblik sheme za opis pravila naziva se *Document Type Definition* (DTD). Budući da je jako rasprostranjen, još uvijek je u širokoj upotrebi. Osim DTD sheme, W3C je razvio i noviji oblik koji se naziva *XML Schema*. Osim ove dvije postoje i druge sheme, međutim dublji ulazak u analizu i usporedbu shema udaljilo bi ovu raspravu od glavne teme.¹⁰

Na kraju poglavlja o svakoj vrsti gradiva tablično će se prikazati kako svaki od formata udovoljava tehničkim zahtjevima za dugoročno očuvanje dokumenata i kako svaki od formata udovoljava zahtjevima kvalitete i funkcionalnosti.

3.1. Formati za tekstualno gradivo

Kriteriji kvalitete i funkcionalnosti koje se promatra prilikom odabira najpovoljnijeg formata za tekstualnu građu (u nju su uključeni i notni zapisi glazbenih djela) su:

Ispravan prikaz tekstualnih dijelova uključuje prikladan linearan tijek čitanja na ekranu, mogućnost ispisa dijelova dokumenta na papir, izvoz navoda, pretraživanje riječi unutar teksta i indeksiranje u svrhu pretraživanja dijelova dokumenta. Prikaz

¹⁰ XML. Wikipedia. URL: <https://en.wikipedia.org/wiki/XML>. (29. lipnja 2018.)

svakog tekstualnog dijela mora odražavati namjeru autora kroz prikaz pojedinačnih slova, strukture odlomaka, lista, naslova i načina naglašavanja.

Integritet strukture dokumenta i kretanje po dokumentu opisuje u kojoj je mjeri funkcija kretanja po dokumentu i automatska analiza omogućena logičkom strukturom dokumenta ključna za njegovu upotrebljivost. Ova mjera vrlo je važna za visoko strukturirane dokumente poput rječnika, enciklopedija i ostalih radova koji imaju logičku strukturu.

Integritet izgleda stranice, vrste pisma i ostalih obilježja dizajna opisuje u kojoj mjeri su izgled i točan odabir značajki poput vrste pisma i izgleda stupaca ključni za njegovo značenje. Važnost ovog kriterija bit će veća ako sam izgled stranice prenosi informaciju, dok će u slučaju dokumenta koji se sastoji isključivo od teksta ovaj kriterij biti od manje važnosti.

Integritet prikaza matematičkih i kemijskih formula, dijagrama i sl. govori u kojoj je mjeri prikaz netekstualnih elemenata dokumenta ključan za njegov informacijski sadržaj.

Kada se govori o formatima za pohranu tekstualne građe, njih možemo svrstati u dvije kategorije. Prva su formati koji koriste tzv. strukturirani označni jezik (engl. *Structural Markup Language*) temeljen na XML jeziku, a druga su formati koji prikazuju izgled stranice. Američka Kongresna knjižnica prednost daje prvom formatu ukoliko je tekst nastao u tom formatu ili je u tom formatu pripremljen za objavljivanje.¹¹

Različite institucije donose različite preporuke u vezi formata za tekstualno gradivo. Sljedeća tablica prikazuje formate koje Kongresna knjižnica preporučuje za svaku vrstu tekstualnog gradiva te preporučene metapodatke.

¹¹ Arms et al., Sustainability of Digital Formats, n. dj.

Tablica 1: Preporučeni formati za tekstualno gradivo

Vrsta tekstualnog gradiva	Preporučeni format	Metapodaci
Digitalni tekstualni radovi	BITS EPUB PDF/UA PDF/A PDF	Naslov Stvaratelj Datum nastanka Mjesto izdavanja Izdavač / proizvođač / raspačavač ISBN Podaci o kontaktu Jezik djela* Ostali relevantni identifikatori (DOI, LCCN, originalni URL i sl.)* Izdanje* Deskriptori teme* Sažeci*
Digitalne serijske publikacije	NISO JATS PDF/UA PDF/A PDF	Naslov serije ISSN i ISSN-L Izdavač Učestalost izlaženja Mjesto izdavanja Svezak Broj Datum izdanja Naslov članka Autor članka Identifikator članka (DOI, originalni URL i sl.) Ostali opisni metapodaci*
Notni zapisi glazbenih djela	MusicXML MEI PDF/UA PDF/A PDF	Naslov Stvaratelj Datum stvaranja Mjesto izdavanja Izdavač / proizvođač / raspačavač ISMN Instrumenti Jezik djela* Ostali relevantni identifikatori (ISBN, DOI, LCCN, originalni URL i sl.)* Izdanje* Deskriptori teme* Događaj* Sažeci*
		* uvrstiti ako to format dozvoljava

Ove formate, iako ne ovim redoslijedom, kao preporučene navode i druge ustanove, kao primjerice slovenski Arhiv družboslovnih podatkov^{12,13} i britanski UK

¹² Priporočeni formati : Arhiv družboslovnih podatkov. 2017. URL: <https://www.adp.fdv.uni-lj.si/deli/postopek/priprava/formati/>. (24. svibnja 2018.)

¹³ Štebe, J.; Vipavc Brvar, I. Analiza stanja in perspektiva digitalne hrambe v arhivu družboslovnih podatkov (ADP). // Knjižnica, 55(2011)1, str. 57-86

Data Service.¹⁴ I drugi znanstvenici koji istražuju područje arhivistike i dugoročnog očuvanja dokumenata ističu PDF/A format kao standard za pohranu tekstualnih dokumenata.¹⁵

3.1.1. BITS format

BITS (engl. *Book Interchange Tag Suite*) format u inačici 2.0 objavljen je u veljači 2016., a sadrži XML model za knjige temeljen na specifikacijama JATS (engl. *Journal Article Tag Suite*) inačice 1.1. Svi formati temeljeni na XML strukturi opisani su kroz definiciju tipa dokumenta (DTD).

Cilj autora ovog formata je ponuditi opći format u kojem izdavači i arhivisti mogu razmjenjivati konačan sadržaj knjiga, uključujući dijelove knjige poput poglavlja. Iako sam format održava i dokumentira Nacionalni centar za biotehnoške informacije (NCBI – *National Center for Biotechnology Information*) Nacionalne medicinske knjižnice (NLM – *National Library of Medicine*) model je upotrebljiv i izvan okvira medicinske struke. Format podržava oznake za metapodatke i narativni sadržaj knjige, metapodatke i narativni sadržaj komponenata knjiga i metapodatke na razini kolekcije za serije knjiga i sveske. Uz tekstualni sadržaj, ovaj format podržava i grafički sadržaj knjiga kao i razmjenu dijelova knjiga.¹⁶ Datoteke ovog tipa imaju nastavak **.xml**.

3.1.2. EPUB format

EPUB (engl. *Electronic PUblication*) format namijenjen je elektroničkim publikacijama s prikazom teksta koji se prilagođava ekranu uređaja na kojem se tekst pregledava, za razliku od formata s unaprijed definiranom paginacijom. Unatoč tome EPUB format nudi izdavačima neke mehanizme za prikaz dokumenata s unaprijed zadanim izgledom stranice. EPUB format ima dvije inačice, inačicu 2 i inačicu 3.

Za razliku od inačice 2, inačica 3 ovog formata omogućuje dodavanje zvučnog i video zapisa, te sinkronizaciju teksta sa zvukom. Ukoliko gradivo sadrži video zapis, izdavač je dužan osigurati i nepokretnu sliku za pregled gradiva na uređajima koji ne podržavaju gledanje video zapisa.

¹⁴ Recommended formats : UK Data Service. URL: <https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats>. (24. svibnja 2018.)

¹⁵ Kavčić-Čolić, A. Trajno ohranjanje digitalnih virov: Koncepti in metode. // Knjižnica, 54(2010)1-2, str 99-119.

¹⁶ Arms et al., Sustainability of Digital Formats, n. dj., URL: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000453.shtml>. (25. svibnja 2018.)

EPUB zapravo označava skup dokumenata koji sadrže samu građu, paketni dokument, barem jedan navigacijski dokument i druge prateće datoteke različitih tipova, uključujući i grafičke datoteke zapakirane u tzv. OCF spremnik (OCF – *Open Container Format*) temeljen na ZIP formatu.

Paketni dokument je XML datoteka koja specificira sve dokumente koji sadrže građu i koja zajedno s pratećim datotekama definira redoslijed čitanja dokumenta.

Navigacijski dokument je datoteka u HTML5/XHTML formatu koja omogućuje navigaciju kroz dokument, bez obzira kreće li se kroz dokument osoba ili stroj.¹⁷ Datoteke ovog tipa imaju nastavak **.epub**.

EPUB format održava Međunarodni forum za digitalno izdavaštvo (IDPF – *International Digital Publishing Forum*).

EPUB format zadovoljava većinu poželjnih osobina formata za tekstualnu građu ukoliko sadržaj nije kriptiran i ukoliko nisu primijenjene druge metode tehnološke zaštite koje sprečavaju dugoročnu zaštitu i pristup.

Nedostatak ovog formata je što metapodaci nisu ugrađeni u skup dokumenata, već su dostupni jedino kroz poveznice prema vanjskim zapisima, što znači da usporedno s pristupom sadržaju treba osigurati i pristup datotekama s metapodacima.

3.1.3. PDF formati

Iako se PDF (engl. *Portable Document Format*) ne nalazi na prvom mjestu formata koji čuvaju izgled stranice, opisan će ga se na početku pregleda formata iz PDF obitelji formata budući da se ostali formati iz ove obitelji temelje na ovom formatu. Dokumenti predstavljeni PDF formatom mogu biti strukturirani ili jednostavni. Oni mogu sadržavati tekst, slike, grafiku i drugi multimedijски sadržaj poput pokretnih slika i zvuka. Format podržava anotacije, metapodatke, hipertekst poveznice i zabilješke (engl. *bookmark*). Metapodaci dokumenta bilježe se kroz XMP (engl. *eXtensible Metadata Platform*) oznake. Kasnije inačice formata mogu imati ugrađene i dodatne funkcionalnost poput geoprostornih informacija unutar dokumenata koji predstavljaju geografske karte ili druge geoprostorne slike poput satelitskih fotografija. Inačica 1.7 iz studenoga 2006. odobrena je kao međunarodna norma ISO 32000-1:2008.¹⁸ Datoteke iz PDF obitelji formata imaju nastavak **.pdf**.

¹⁷ Ibid. URL: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000308.shtml>. (25. svibnja 2018.)

¹⁸ Ibid. URL: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000030.shtml>. (27. svibnja 2018.)

3.1.3.1 PDF/A format

Format PDF/A član je obitelji ISO 19005 norme namijenjen dugoročnom očuvanju dokumenata s očuvanjem izgleda stranice.

Inačica PDF/A-1 temelji se na formatu PDF/A, inačica 1.4, a objavljena je 2005. godine kao norma ISO 19005-1. Inačica PDF/A-2 iz 2011., objavljena kao norma ISO 19005-2 proširuje mogućnosti formata PDF/A-1 i utemeljena je na formatu PDF/A, inačici 1.7. Posljednja inačica, PDF/A-3 u odnosu na svoju prethodnicu uvodi samo jednu novinu – u dokument je omogućena ugradnja datoteka bilo kojeg formata.

Ograničenja koja se u odnosu na PDF format postavljaju pred PDF/A usklađeni dokument su:

- Zabrana zvukovnih i video sadržaja
- Zabrana Javascript programa i pokretanja izvršnih datoteka
- Svi obrasci moraju biti ugrađeni u dokument i s pravnom dozvolom za ugradnju i prikaz
- Boje moraju biti opisane neovisno o uređaju na kojem se dokument čita
- Zabrana šifriranja sadržaja
- Obveza upotrebe metapodataka definiranih normom.

Ovaj format ima tri razine usklađenosti. Razina A zadovoljava sve kriterije specifikacije. Za datoteku ovog stupnja usklađenosti kaže se da je PDF/A-1a usklađena. Razina usklađenosti B u odnosu na razinu A ne mora podržavati preslikavanje tekstualnih znakova na UNICODE¹⁹ znakove i ne mora odražavati prirodni tijek čitanja dokumenta. Za datoteku ovog stupnja usklađenosti kaže se da je PDF/A-1b usklađena.²⁰ Dokumenti razine usklađenosti U imaju veći stupanj usklađenosti od dokumenata PDF/A-1b utoliko što se cijeli tekst mora moći preslikati na UNICODE znakove, što osigurava da ga se može cijeloga indeksirati i prikazati.²¹

3.1.3.2 PDF/UA format

PDF/UA format temelji se na PDF formatu i svrha mu je osiguranje dostupnosti i podrška za tehnologije koje pomažu korisnicima s posebnim potrebama. Skraćenica

¹⁹ UNICODE je standard u računalnoj industriji koji omogućuje konzistentno kodiranje i prikaz tekstualnih znakova u većini svjetskih pisama.

²⁰ Document management – Electronic document file format for long-term preservation : Part1: Use of PDF 1.4 (PDF/A-1) : 2005. Geneva : International Organization for Standardization.

²¹ Arms et al., Sustainability of Digital Formats, n. dj. URL: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000321.shtml>; (30. svibnja 2018.)

UA označava opću dostupnost (engl. *Universal Accessibility*) iako se ovaj izraz ne spominje u normi ISO 14289-1 koja definira ovaj format. Kako bi PDF format bio sukladan sa zahtjevima za PDF/UA, mora, između ostalih, zadovoljiti sljedeće kriterije:

- Cjelokupni sadržaj mora biti označen, dok njegovi pomoćni dijelovi, tzv. artefakti ne smiju.
- Sadržaj treba označiti u strukturnom stablu semantički prikladnim oznaka u logičnom slijedu čitanja. Norma nalaže detaljne propise za uporabu naslova, tablica i popisa kako bi se moglo podržati tehnologiju za pomoć korisnicima s posebnim potrebama.
- Svi strukturirani tipovi koji se koriste trebaju biti standardni ili se preslikati na standardne tipove.
- Grafički objekti (slike) koje tvore stvarni sadržaj treba označiti oznakom Slika, koja uključuje alternativni ili zamjenski tekst; opisi uz slike moraju se označiti kao opisi.
- Informaciju se ne smije prenositi kroz kontrast, boju, format ili izgled stranice, osim ako sadržaj nije označen tako da se opisuje njegovo značenje.
- Sve informacije koje se prenose zvukom moraju biti dostupne i bez korištenja zvuka.
- Pisma (fontovi) moraju biti ugrađena.
- Nisu dopušteni dinamički obrasci, dok statički jesu.

Ukoliko je PDF/UA dokument ujedno usklađen s PDF/A zahtjevima to ga stavlja na vrh popisa preporučenih formata za dugoročno očuvanje tekstualnih dokumenata koji čuvaju izgled stranice.²²

3.1.3.3 PDF/X format

PDF/X format je dio PDF obitelji formata namijenjen razmjeni grafičkih datoteka i pripremi za tisak. Ovaj format zabranjuje sve elemente PDF specifikacije koji ne služe izravno pripremi za tisak. Zbog svoje specifičnosti ovaj format ne razmatra se kao preporučeni za dugoročno očuvanje gradiva, ali ga valja spomenuti kako bi se zaokružio pregled PDF obitelji formata.²³

²² Ibid. URL: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000350.shtml>. (27. svibnja 2018.)

²³ Ibid. URL: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000124.shtml>. (29. svibnja 2018.)

3.1.4. NISO/JATS format

ANSI/NISO Z39.96 je specifikacija za Standardizirane oznake za članke u časopisima, poznata i pod nazivom JATS (engl. *Journal Article Tag Suite*). JATS se temelji na sustavu oznaka koji je razvila Nacionalna medicinska knjižnica (NLM), članica Nacionalnog centra za biotehno­loške informacije (NCBI). Od 2009. godine JATS se razvija u okviru organizacije NISO (*National Information Standards Organization*).

JATS ima tri inačice. Zelena inačica za arhiviranje i razmjenu definira elemente i attribute koji opisuju sadržaj i metapodatke članka, uključujući istraživačke i neistraživačke članke, pisma, uvodnike i recenzije knjiga i proizvoda. Ova inačica omogućuje opis cijelog sadržaja članka ili samo metapodataka. Također omogućuje očuvanje slijeda sadržaja i generiranog teksta.

Plava inačica optimizirana je za arhive koji žele regulirati i kontrolirati sadržaj umjesto prihvatiti slijed i raspored koji je odredio izdavač. Ova inačica također je namijenjena za početno XML označavanje časopisa, najčešće prilikom konverzije s izvornog formata poput Microsoft Worda.

Narančasta inačica optimizirana je za autorstvo novih članaka, gdje je reguliranost i kontrola sadržaja važna i gdje je korisnije imati samo jedan način označavanja strukture. Ova inačica sadrži mnoge elemente čiji sadržaj se mora javljati u točno određenom redoslijedu. S obzirom da je ova inačica namijenjena za korištenje prije prihvaćanja članka, ona ne sadrži metapodatke za opis samog članka.²⁴ Datoteke u NISO/JATS formatu imaju nastavak **.xml**.

3.1.5. MusicXML format

MusicXML format osmišljen je za prikaz notnih zapisa, osobito onih pisanih uobičajenom zapadnom glazbenom notacijom od 17. stoljeća nadalje. Koristi se kao format za razmjenu između aplikacija za stvaranje, analizu, dohvat i izvođenje notnih zapisa. MusicXML utemeljen je na osnovnim obilježjima MIDI (engl. *Musical Instrument Digital Interface*)²⁵ formata, ali je prikaz elemenata izravan i čitljiv čovjeku.

²⁴ Ibid. URL: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000451.shtml>. (31. svibnja 2018.)

²⁵ MIDI je tehnički standard koji opisuje komunikacijske protokole, digitalno sučelje i električne veze između širokog spektra elektroničkih glazbenih instrumenata, računala i drugih glazbenih i audio uređaja.

Ovaj format može prenijeti dovoljno informacija za ispis notnog zapisa, ali sam po sebi, za razliku od PDF formata, ne prikazuje note.

Ovaj format može bilježiti notni zapis na dva načina: po vremenu (*time-wise*) ili po glazbenim dionicama (*part-wise*). Ako je zapis načinjen po dionicama, datoteka je strukturirana tako da djelo sadrži dionice, svaka dionica taktove, a svaki takt note. U vremenskom zapisu takt i dionica zamijenjeni su u hijerarhiji.²⁶ Ovisno o načinu bilježenja u datoteci se koristi glavni element <score-partwise> za zapis po dionicama s taktovima u svakom dijelu, odnosno glavni element <score-timewise> za vremenski zapis gdje su u svakom taktu opisane note za svaku od dionica.

Element <opus> koristi se kao glavni element ako se želi spojiti više dijelova u jedno djelo, na primjer stavke u simfoniji.

Datoteke nastale u ovom formatu često su velike pa se uglavnom isporučuju sažete.²⁷ Datoteke u MusicXML formatu imaju nastavak **.xml**.

3.1.6. MEI format

MEI (engl. *Music Encoding Initiative*) je inicijativa koja nastoji definirati sustav bilježenja glazbenih dokumenata za strojno čitanje. Cilj stručnjaka okupljenih oko ove inicijative je uspostavljanje najbolje prakse za predstavljanje širokog spektra glazbenih dokumenata i struktura. Rezultat ovih nastojanja je MEI shema, osnovni skup pravila za bilježenje fizičkih i intelektualnih obilježja notnih zapisa u XML obliku. Ovaj osnovni skup pravila dopunjen je MEI smjernicama s detaljnim objašnjenjima elemenata MEI modela i prijedlozima za najbolji način korištenja.

MEI format razvija i održava Akademie der Wissenschaften und der Literatur iz Mainza.

Za razliku od MusicXML formata koji je prvenstveno namijenjen razmjeni između aplikacija za obradu notnih zapisa, MEI format uz te funkcionalnosti sadrži i informacije o notaciji i njezinom intelektualnom sadržaju u strukturiranom i sistematičnom obliku. Osim toga MEI ne podržava samo zapadni notacijski sustav već i srednjovjekovnu neumatsku notaciju i renesansnu menzuralnu notaciju. Neumatska notacija je najraniji način zapisivanja nota koji nije određivao njihovo trajanje.

²⁶ Bello, J. P.; Chew, E.; Trunbull, D. ISMIR 2008: Proceedings of the 9th International Conference of Music. Philadelphia : Drexel University, dostupno i na: <https://books.google.hr/books?id=OHp3sRnZD-oC>. (31. svibnja 2018.)

²⁷ Arms et al., Sustainability of Digital Formats, n. dj. URL: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000358.shtml>. (31. svibnja 2018.)

Menzuralna notacija uvodi bilježenje trajanja pojedinih tonova.²⁸ Ove notacije MEI podržava ne samo kroz vizualni prikaz, već i zadržavajući strukturu i semantiku notacije kako bi zapis točno prikazao.²⁹ Datoteke u MEI formatu imaju nastavak **.xml**.

²⁸ Andreis, J. Povijest glazbe. 1975. Zagreb : Liber Mladost.

²⁹ An introduction to MEI. Music Encoding Initiative. URL: <http://music-encoding.org/about/>. (3. lipnja 2018.)

Tablica 2: Usporedba formata za pohranu tekstualne građe prema tehničkim kriterijima

Format	Otkrivenost	Usvojenost	Transparentnost	Samodokumentiranje	Vanjske ovisnosti	Utjecaj patenata	Mehanizmi tehničke zaštite
BITS	Dokumentira ga NCBI/NLM i javno je dostupan.	Inačica 2.0 je relativno novi format nastao 2017. Ipak, danas mnoge institucije prihvaćaju ovaj format (Portico, The Publication Office for the EU...)	Visoka transparentnost zbog korištenja XML strukture. Sadržaj se može pročitati osnovnim alatima za uređivanje teksta.	DTD sadrži veliki skup metapodataka	Nema.	Nema. Specifikacija formata je u javnoj domeni.	Nema.
EPUB	Otvoreni standard koje se održava i razvija pod pokroviteljstvom IDPF-a. IDPF također brine o dokumentaciji.	Inačica 3 je relativno novi format nastao 2012. i tek se počinje usvajati. Dostupan je besplatan validator. Također je dostupan plug-in za Chrome Internet preglednik. Adobe InDesing već podržava neke od funkcionalnosti ovog formata.	Tekst mora biti u XHTML/HTML5 formatu što ga čini vrlo transparentnim. Budući da je dozvoljeno šifriranje, paket mora sadržavati informacije potrebne za dešifriranje, uključujući i algoritam. Ugrađeni fontovi mogu biti skriveni djelomičnom enkripcijom pa i to može smanjiti transparentnost.	Obavezni elementi su naslov, identifikator i jezik, kao i indikator datuma promjene. Dozvoljena je poveznica na metapodatke koji nisu sadržani u paketu. Paketni dokument također sadrži manifest komponenta.	Za sadržaj koji nije kriptiran nema ovisnosti. U slučaju da je sadržaj šifriran i/ili je dostup ograničen putem DRM funkcionalnosti (DRM – <i>Digital Rights Management</i>) čitljivost sadržaja ovisi o tome ima li čitač ugrađene odgovarajuće aplikacije za dešifriranje i/ili DRM.	Nema.	Osim mogućnosti enkripcije, i djelomične enkripcije fontova, dodatni element zaštite je upravljanje digitalnim pravima.

Format	Otkrivenost	Usvojenost	Transparentnost	Samodokumentiranje	Vanjske ovisnosti	Utjecaj patenata	Mehanizmi tehničke zaštite
PDF/UA	Otvorena norma koju je objavio ISO u srpnju 2012.	Od objavljivanja norme nekoliko proizvođača PDF čitača uključujući Adobe i NVDA (<i>Non-Visual Desktop Access</i>) obvezalo se podržavati ovu normu.	Ovisi o tome da li je program za čitanje usklađen sa specifikacijama. Alati za izradu dokumenta su sofisticirani. Format PDF/UA dozvoljava šifriranje. Potrebno je aktivirati poseban ključ kako bi se šifriranom sadržaju omogućio pristup putem tehnologije za pomoć korisnicima s posebnim potrebama.	Metapodaci moraju biti pohranjeni u XMP formatu.	Sva pisma (fontovi) moraju biti ugrađena u dokument.	Brojni patenti pokrivaju tehnologiju u pozadini PDF specifikacije, počevši od inačice 1.3 nadalje. Patenti se smiju koristiti za razvoj aplikacija koje stvaraju PDF dokumente, ali aplikacije koje ih interpretiraju su izuzete iz ove dozvole.	Format PDF/UA dozvoljava šifriranje. Potrebno je aktivirati poseban ključ kako bi se šifriranom sadržaju omogućio pristup putem tehnologije za pomoć korisnicima posebnim potrebama.
PDF/A	Obitelj otvorenih normi. Ova obitelj formata definirana je sljedećim dokumentima: Norma ISO 19005-1 i dokument PDF Reference, Third Edition, Version 1.4 za format PDF/A-1; Norma ISO 19005-2 i norma ISO 32000-1 za format PDF/A-2; Norma ISO 19005-3 i norma ISO 32000-1 za format PDF/A-3	Ovaj format široko je usvojen. Podržava ga Adobe Acrobat Professional počevši od inačice 7, zatim Microsoft Office počevši od inačice 2007, slijedi Open Office počevši od inačice 2.4. Mnoge komercijalne tvrtke nude proizvode koji podržavaju stvaranje, migraciju i validaciju PDF/A datoteka. Među njima su: Apago, Inc., Visioneer, Callas	Ovisi o tome da li je program za čitanje usklađen sa specifikacijama. Alati za izradu dokumenta su sofisticirani.	Metapodaci moraju biti pohranjeni u XMP formatu.	Format PDF/A je ograničen u odnosu na osnovni PDF format u cilju izbjegavanja vanjskih ovisnosti. Sva pisma moraju biti ugrađena u dokument.	Brojni patenti pokrivaju tehnologiju u pozadini PDF specifikacije, počevši od inačice 1.3 nadalje. Patenti se smiju koristiti za razvoj aplikacija koje stvaraju PDF dokumente, ali aplikacije koje ih interpretiraju su izuzete iz ove dozvole. Uz usvajanje formata PDF, inačice 1.7 kao ISO norme (ISO 32000-1), Adobe je dozvolio svim pojedincima i institucijama na svijetu	Format PDF/A ne dozvoljava šifriranje.

Format	Otkrivenost	Usvojenost	Transparentnost	Samodokumentiranje	Vanjske ovisnosti	Utjecaj patenata	Mehanizmi tehničke zaštite
		Software i drugi. Neke aplikacije u slobodnom pristupu kao što je FOP podržavaju minimalni PDF/A-1b format. Sve veći broj knjižnica i arhiva preporuča ili zahtijeva korištenje formata PDF/A. Budući da je postizanje usklađenosti s PDF/A-1 formatom teško ostvarivo, prihvatljivim se smatra i usklađenost s formatom PDF/A-1b. U SAD-u Vlada sve više podupire uporabu formata PDF/A.				besplatno pravo upotrebe formata PDF/A.	
PDF	U cijelosti dokumentiran. Razvija ga Adobe Systems Incorporated, a specifikacije su dostupne i besplatne.	Izrazito usvojen kao format za dijeljenje dokumenata s unaprijed određenim izgledom stranice neovisan o platformi. Program za čitanje, Adobe Reader besplatan je i dostupan za gotovo sve platforme koje su danas u upotrebi.	Ovisi o tome da li je program za čitanje usklađen sa specifikacijama. Alati za izradu dokumenta su sofisticirani.	Novije inačice PDF formata mogu uključiti XMP metapodatkovne pakete.	Vjieran prikaz dokumenta zahtijeva da pisma budu ugrađena u dokument.	Brojni patenti pokrivaju tehnologiju u pozadini PDF specifikacije, počevši od inačice 1.3 nadalje. Patenti se smiju koristiti za razvoj aplikacija koje stvaraju PDF dokumente, ali aplikacije koje ih interpretiraju su izuzete iz ove dozvole.	PDF format nudi nekoliko oblika tehničke zaštite. Oni obuhvaćaju i šifriranje, što ne osigurava dostupnost u buduću tehnološkim okruženjima.
NISO JATS	Otvoreno dokumentiran i besplatan za	Počevši od 2005. godine Portico je surađivao s	Izrazito transparentan format. Sadržaj teksta je	Format sadrži bogat skup elemenata za	Nema.	Nema.	Nema.

Format	Otkrivenost	Usvojenost	Transparentnost	Samodokumentiranje	Vanjske ovisnosti	Utjecaj patenata	Mehanizmi tehničke zaštite
	preuzimanje. Svi elementi su javno dostupni.	knjižnicama i izdavačima na očuvanju e-časopisa i aktivno razvijao i održavao JATS. Članci s punim tekstom dostavljeni u SGML ili XML formatu konvertirani su u JATS. Južnoamerička mreža SciELO, koja posluje od 1988. u 15 južnoameričkih država i u Južnoafričkoj Republici., u travnju 2014 najavila je da prelazi na upotrebu XML DTD-a temeljenog na JATS specifikacijama. ISO koristi za objavu svojih normi prilagođeni JATS format koji se naziva ISOSTS (<i>ISO Standards Tag Set</i>). Brojni softverski alati prilagođeni su za pripremu članaka prema JATS specifikacijama. To su na primjer eXtyles NLM – dodatak za Microsoft Word koji podržava pripremu i	u XML formatu i kao takav čitljiv je u osnovnim alatima za obradu teksta, Internet preglednicima i sl. Elementi imaju razumljive oznake, a dijelovi dokumenta raspoređeni su u prirodnom slijedu čitanja.	metapodatke na razini članaka i časopisa.			

Format	Otkrivenost	Usvojenost	Transparentnost	Samodokumentiranje	Vanjske ovisnosti	Utjecaj patenata	Mehanizmi tehničke zaštite
		izvoz članaka u JATS formatu, zatim JATSkIt, custom Jats writer for Pandoc, JATS Preview Stylesheets i PubReader.					
MusicXML	U potpunosti otkriven.	Više od 230 programa za notne zapise u nekoj mjeri podržava MusicXML, uključujući većinu programa za pisanje nota i većinu glazbenih OCR programa i većina programa za sekvencioniranje glazbe. ³⁰	Vrlo transparentan.	XML DTD uključuje identifikacijski dio s elementima u zaglavlju u formatu MuseData i Dublin Core. Različiti elementi mogu sadržavati bilo koji opisni metapodatak koji se odnosi na zapis kao cjelinu.	Nema.	Slobodan za upotrebu.	Nema.
MEI	U potpunosti otkriven i dokumentiran	Danska kraljevska knjižnica za zapis bibliografskih informacija o djelima Carla Nielsena. Beethoven Digital Edition. ³¹ Za sada format nije u širokoj upotrebi.	Vrlo transparentan.	MEI shema uključuje brojne metapodatke o glazbenom djelu.	Nema.	Slobodan za upotrebu.	Nema.

³⁰ MusicXML. Wikipedia. URL: <https://en.wikipedia.org/wiki/MusicXML>. (31. svibnja 2018.)

³¹ ahankinson. Music Encoding Initiative. 5. veljače 2015. <https://musescore.org/en/node/24208>. (3. lipnja 2018.)

Tablica 3: Usporedba preporučenih formata za pohranu tekstualne građe prema kriterijima kvalitete i funkcionalnosti

Format	Ispravan prikaz tekstualnih dijelova	Integritet strukture dokumenta i kretanje po njemu	Integritet izgleda stranice, vrste pisma i ostalih obilježja dizajna	Integritet prikaza matematičkih i kemijskih formula, dijagrama i sl.
BITS	Izvršna podrška.	Dokumentiranje logičke strukture knjige, uključujući i njen odnos prema seriji knjiga ključna je funkcionalnost ovog formata.	Mogućnost reproduciranja točnog formata knjige nije svrha ovog formata.	Moguće je uključiti MathML i TeX math DTD, ali prikaz može biti ograničen zbog ograničenja koja ima MathML.
EPUB	Dobra podrška.	Očuvanje logičke strukture dokumenta ključna je funkcionalnost ovog formata.	Izdavači donekle mogu upravljati nekim aspektima izgleda stranice kroz stranice stila (engl. <i>style-sheets</i>), ali kod prikaza teksta koji se prilagođava ekranu čitača reci teksta prelamat će se na različitim mjestima kad se čitaju na različitim platformama i pri različitim korisničkim postavkama.	Moguće je uključiti MathML DTD.
PDF/UA	Dobro podržava. Znakovi teksta moraju se preslikati na UNICODE, a tekst mora biti dostupan u logičnom slijedu. Dokumenti nastali skeniranjem stranica zahtijevaju poveznicu s tekstem. Tekst nastao optičkim prepoznavanjem znakova mora se ispraviti i označiti kako bi prikazao strukturu dokumenta.	Mnoga postavljena ograničenja odnose se na obavezno i pravilno korištenje strukturnih oznaka.	PDF format osmišljen je tako da očuva izgled stranice. Ipak, zahtjevi za dostupnost mogu imati za posljedicu da se PDF/UA dokument ne podudara u potpunosti s izgledom izvornika.	Slika koja predstavlja formulu mora se opisati tekstem.
PDF/A	Moguće je postići dobar rezultat, ali on nije zajamčen. Format PDF/A ne sprečava stvaranje dokumenata iz skeniranih slika, ali takve datoteke ne moraju podržavati indeksiranje teksta ili izvoz citata.	Logička struktura dokumenta održana je u PDF/A datoteci jedino ako je stvaratelj dokumenta proveo strukturno označavanje.	PDF format osmišljen je tako da očuva izgled stranice.	Mogu se prikazati kao ugrađene slike.

Format	Ispravan prikaz tekstualnih dijelova	Integritet strukture dokumenta i kretanje po njemu	Integritet izgleda stranice, vrste pisma i ostalih obilježja dizajna	Integritet prikaza matematičkih i kemijskih formula, dijagrama i sl.
PDF	PDF format namijenjen je za autentičan prikaz stranice. Ispis i uvećanje funkcionalnosti su čitača. Kvaliteta rasterskih slika ovisi o kvaliteti same ugrađene slike. Dobra podrška je moguća, ali nije zajamčena. Stvarateljima dokumenata moguće je zabraniti ispis i/ili izvoz teksta u svrhu citiranja.	Logička struktura dokumenta se prikazuje samo ako autor ili proces koji kreira dokument poduzme korake koji bi uključili strukturalne oznake tijekom izrade.	PDF format osmišljen je tako da očuva izgled stranice.	Mogu se prikazati kao ugrađene slike.
NISO JATS	Dobra podrška.	Očuvanje logičke strukture dokumenta ključno je obilježje JATS formata.	Namjera je očuvati intelektualni sadržaj časopisa neovisno o tome u kojem je obliku sadržaj dostavljen.	Moguće je uključiti MathML i TeX math DTD, ali prikaz može biti ograničen zbog ograničenja koja ima MathML.
MusicXML	Iako ovo nije format namijenjen tekstu, formati za notne zapise dijele funkcionalnosti uobičajene za tekst. MusicXML može se koristiti za prikaz i ispis notnih zapisa. Tekstualni dijelovi kao što su stihovi mogu se pretraživati putem alata za indeksiranje teksta.	MusicXML predstavlja semantičku strukturu notnog zapisa: tonalitet, taktove, note s naglascima i trajanjem, glazbene dionice za različite instrumente te riječi/stihove.	MusicXML ne definira prijelome stranica i sl. On predstavlja elemente izgleda stranice važne za ispis nota poput toga hoće li oznake biti iznad ili ispod note, veličinu elemenata ispisa u odnosu na takt koja se može prilagoditi određenoj veličini stranice.	Ne podržava.
MEI	Uz pomoć pomoćnog programa koji sadržaj MEI formata transformira u SVG (engl. <i>Scalable Vector Graphic</i>) format koji se dalje koristi za ispis. ³²	MEI zapis predstavlja semantičku strukturu notnog zapisa: tonalitet, taktove, note s naglascima i trajanjem, glazbene dionice za različite instrumente te riječi/stihove.	MEI zapis ne definira prijelome stranica i sl. Izgled stranice može se podesiti u okviru pomoćnog programa.	Ne podržava.

³² A music notation engraving library. URL: <http://www.verovio.org/index.xhtml>. (3. lipnja 2018.)

3.2. Formati za nepokretne slike

Nepokretnim slikama smatramo onu građu koja svoje značenje prenosi na vizualan način, primjerice fotografije, postere, grafikone, dijagrame i sl. Ova građa može se nalaziti u dvije glavne skupine datoteka – vektorske i rasterske. Vektorske slike sastoje se od objekata – crta, krivulja i likova – koji se opisuju matematički i može ih se smanjivati i uvećavati bez gubitka kvalitete. Rasterske slike sastoje se od polja točaka, tzv. piksela i najčešće su rezultat skeniranja ili fotografiranja digitalnim kamerama. Uvećavanjem rasterske slike gube kvalitetu, tj. postaju zrnate.

Prilikom odabira najpovoljnijih formata za dugotrajno očuvanje građe u obliku nepokretnih slika valja voditi računa o sljedećim kriterijima kvalitete i funkcionalnosti.

Ispravan prikaz nepokretnih slika odnosi se na prikaz na zaslonu i na ispis na pisaču. Neki korisnici žele imati mogućnost uvećavanja slike radi proučavanja detalja, a drugi mogućnost prikaza na uređajima različite veličine i razlučivosti zaslona. Nekim korisnicima važan je ispis u visokoj kvaliteti.

Jasnoća je značajka koja se odnosi na rasterske slike. Ona govori o tome u kojoj mjeri format može predstaviti sadržaj visoke razlučivosti. Dvije veličine najčešće se povezuju s pojmom jasnoće. To su broj točaka po jedinici duljine (npr. točke po inču) i bitna dubina. Bitna dubina je podatak koji govori s koliko bitova je opisana svaka od točaka slike. Veća bitna dubina i veći broj točaka po jedinici duljine rezultira većom jasnoćom. S druge strane, sažimanje datoteke uz gubitke doprinosi smanjenju jasnoće.

Iako znatno manje, vektorske slike također mogu biti podložne gubitku jasnoće. Kod nekih vektorskih formata moguće je odrediti želi li se prilikom uvećanja slike sačuvati jasnoća rubova, što može dovesti do neznatnog premještanja crta, ili se želi sačuvati geometrijska preciznost, što može narušiti jasnoću rubova. Također, ukoliko je slika kombinacija vektorske grafike s ugrađenim rasterskim elementima, tada pri velikom povećanju rasterski elementi mogu postati nejasni.

Neke se slike sastoje od nekoliko slojeva. Ponekad se ti slojevi prije konačne pohrane spoje u jedan, a ponekad je potrebno zadržati slojeve razdvojenima, kao u slučaju izrade ploča za tisak. U tom slučaju jasnoća ovisi o očuvanju strukture slojeva, tj. očuvanju njihovog prostornog položaja.

Očuvanje boje značajka je koja upućuje na to kako format upravlja bojom, odnosno kako obrađuje i transformira podatke o slici kako bi postigao najbolje održanje boje u smislu ljudskog doživljaja boje. U kojoj će se mjeri boja održati ovisi o

možnosti kodiranja boje u različitim prostorima boja i o mogućnosti pohrane metapodataka o sustavima upravljanja bojom.

Podrška za grafičke efekte i tipografiju vezuje se uz formate koji podupiru grafiku u vektorskom obliku. Ova mjera pokazuje koliko neki format podržava upotrebu sjenčanja, filtara i drugih efekata, koliko ima razina transparentnosti i kako upravlja specifikacijom pisama i uzoraka.

Podrška za višestruke spektralne pojaseve ukazuje na to kako format može uključiti višestruke spektralne pojaseve u cilju podrške znanstvenim analizama. Slike snimane u drugim spektrima osim onoga vidljivoga ljudskom oku, sadrže informacije detektirane u specifičnim frekvencijskim pojasevima elektromagnetskog spektra. Budući da su ti drugi spektri nevidljivi format u svojim metapodacima mora dokumentirati na koji dio vidljivog spektra će se preslikati zabilježena informacija iz nevidljivog spektra.³³

U tablici su navedeni formati za slikovno gradivo koje preporuča Kongresna knjižnica. Iz literature je vidljivo da istraživanja prihvatljivih formata za dugoročno očuvanje slikovnog gradiva ukazuju na sličan poredak preporučenih formata.³⁴

Tablica 4: Preporučeni formati za nepokretne slike

Vrsta slikovnog gradiva	Preporučeni format	Metapodaci
Digitalne fotografije	TIFF JPEG2000 PNG JPEG/JFIF DNG BMP GIF	Naslov Stvaratelj Datum nastanka Mjesto izdavanja Izdavač/producent/distributer Informacije o kontaktu Uobičajena ugrađena shema (IPTC, FGDC, ISO 19115)* Jezik djela* Drugi relevantni identifikatori (PLUS ID, DOI, LCCN i sl.)* Deskriptori teme* Sažeci* Reference na svako podatkovno polje i tehničke informacije (npr. EXIF metapodaci iz digitalne kamere)*

³³ Arms et al., Sustainability of Digital Formats, n. dj. URL: https://www.loc.gov/preservation/digital/formats/content/still_quality.shtml. (20. lipnja 2018.)

³⁴ Van Horik, R. Permanent pixels : Building blocks for the longevity of digital surrogates of historical photographs. Doctoral thesis. Delft : Technische Universiteit Delft, 2005.

Vrsta slikovnog gradiva	Preporučeni format	Metapodaci
Ostala digitalna grafička građa	Rasterski formati: TIFF JPEG2000 PNG JPEG/JFIF DNG BMP GIF Vektorski formati: SVG DXF EPS Shapefile Kartografska građa: Najpotpuniji podaci, čak i ako je format zaštićen intelektualnim pravima GeoTIFF GeoPDF GeoJPEG2000 Shapefile	Naslov Stvaratelj Datum nastanka Mjesto izdavanja Izdavač/producent/distributer Informacije o kontaktu Uobičajena ugrađena shema (FGDC, ISO 19115)* Jezik djela* Drugi relevantni identifikatori (DOI, LCCN i sl.)* Deskriptori teme* Sažeci* Reference na svako podatkovno polje i tehničke informacije (npr. EXIF metapodaci iz digitalne kamere)*
		* uvrstiti ako to format dozvoljava

3.2.1. TIFF format

Skraćenica TIFF označava *Tag(ged) Image File Format*. Datoteke u ovom formatu organizirane su u tri dijela: zaglavlja datoteke (IFH – engl. *Image File Header*), zatim direktorija datoteke (IFD – engl. *Image File Directory*) te samih slikovnih podataka. TIFF format teoretski može biti u potpunosti bez trećeg dijela, a može sadržavati i više slika u istoj datoteci. U tom slučaju svaka slika ima svoj direktorij (IFD) i svoj skup slikovnih podataka, dok je zaglavlje datoteke zajedničko. Slike sadržane u TIFF datoteci mogu biti kodirane različitim algoritmima i različitim stupnjevima sažimanja, kao i različitim reprezentacijama boja. Zbog toga ovaj format nazivamo i tzv. omotačem (engl. *wrapper*). Uz standardni TIFF format razvijene su i različite inačice za različite posebne potrebe.

Tako na primjer, TIFF_G4 format sadrži rastersku sliku sažetu korištenjem ITU_G4 sheme namijenjene faks uređajima. Takva slika sastoji se od samo dva tona – crnog i bijelog te je prikladna za prikaz teksta koji ne sadrži nikakvu grafiku.

Piramidalno kodirani TIFF sadrži niz istih rasterskih slika od kojih je svaka sve veće razlučivosti. Pogodan je za upotrebu u kojoj se korisnik više usredotočuje na detalje slike nego na sliku u cijelosti.

TIFF_UNC sadrži nesažetu rastersku sliku i neke mu institucije daju prednost pred ostalim formatima. Ovom formatu moguće je dodati i EXIF (engl. *EXchangeable Image File format for digital still cameras*) metapodatke.

TIFF/IT format namijenjen je pripremi slika za tisak i omogućuje pohranu različito kodiranih rasterskih slika kao i slojeva odvojenih boja koji se kombiniraju prilikom tiskanja.

BigTIFF format nadilazi ograničenje datoteke standardnog TIFF formata koje je zbog 32 bitnog načina adresiranja podataka ograničen na 4 GB. BigTIFF koristi 64 bita pa mu je teoretska maksimalna veličina datoteke oko 18.000 PB.

Datoteke ovog tipa imaju nastavak **.tif** ili **.tiff**.³⁵

3.2.2. JPEG2000 format

JPEG2000 format je format je omotač koji sadrži nepokretne slike kodirane u skladu s JPEG 2000 standardom za sažimanje slika. Standard je osmislila skupina Joint Photographic Experts Group, odakle i dolazi naziv formata. Sažimanje ovim standardom može biti sa i bez gubitaka informacije. U usporedbi s formatom JPEG, ovaj format pruža dodatne mogućnosti kao što su dodavanje kanala za prozirnost, informacija o profilu boje i drugih metapodataka. Datoteke ovog formata imaju nastavak **.j2c** ili **.jp2**.³⁶

3.2.3. PNG format

PNG (engl. *Portable Network Graphics*) definira podatkovni tok i format datoteka za rastersko slikovno gradivo. Sadržaj je prenosiv između različitih platformi i sažet bez gubitaka. Boje mogu biti kodirane kao indeksirane, kao nijanse sive i kao RGB boja uz kanal za prozirnost kao neobavezan. Format također sadrži i podatke za korekciju boja za njihov vjeran prikaz na različitim platformama. Datoteke u ovom formatu imaju nastavak **.png**.³⁷

³⁵ Arms et al., Sustainability of Digital Formats, n. dj. URL:

<https://www.loc.gov/preservation/digital/formats/fdd/fdd000022.shtml>. (21. lipnja 2018.)

³⁶ Ibid. URL: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000143.shtml>. (21. lipnja 2018.)

³⁷ Ibid. URL: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000153.shtml>. (21. lipnja 2018.)

3.2.4. JPEG/JFIF format

JFIF (engl. *JPEG File Interchange Format*) je format koji omogućuje razmjenu slika između različitih platformi. On ne uključuje oznake u zaglavlju slike. Ovaj format dozvoljava različite vrste sažimanja sadržaja uključujući i one bez gubitaka.³⁸ Prostor boje u JPEG/JFIF formatu definiran je kroz tri komponente: luminantu Y, te dvije krominantne komponente C_B i C_R koje predstavljaju razliku luminante Y i plave odnosno crvene komponente boje. Zanimljivo je to da svaka od ovih komponenti može biti kodirana u različitoj razlučivosti. Najčešće se Y komponenta koja odražava sivu skalu kodira u većoj razlučivosti dok se komponente boje kodiraju u nižoj razlučivosti.³⁹ Datoteke ovog formata imaju nastavak **.jpg**, **.jpeg**, **.jpe**, **.jfi**, **.jfif**, **.j**, **.jif**, **.jmh** ili **.jfs**.

3.2.5. DNG format

Adobe Digital Negative (DNG) je format za pohranu i razmjenu neobrađenih (*raw*) slika s digitalne kamere. Ovakve slike obično su popraćene i JPEG verzijom. Ovaj format osmišljen je kako bi se profesionalnim fotografima dala veća umjetnička sloboda budući da im je omogućeno manipulirati podacima dobivenima izravno s objektiva kamere. Ovi neobrađeni podaci odgovaraju negativu fotografije u klasičnoj fotografskoj tehnici i zato se često nazivaju digitalni negativ. Takvu neobrađenu sliku se uz pomoć aplikacija smješta u DNG omotač u kojemu se ona najčešće nalazi sažeta bez gubitaka kao JPEG. Aplikacije za pohranu slika u DNG formatu dodaju i skup metapodataka koji su zapravo prošireni skup TIFF metapodataka. Prednost manipuliranja neobrađenim podacima je u tome što korisnik može kroz aplikaciju precizno odrediti niz parametara slike kao što su kontrast, svjetlina, uklanjanje šuma i sl. Ali, za razliku od TIFF ili JPEG formata, neobrađeni podaci ne mogu se koristiti prije obrade u aplikaciji.⁴⁰ Datoteke ovog formata imaju nastavak **.dng** ili **.tif**.

3.2.6. BMP format

Microsoft Windows Bitmap (BMP) format je jednostavan format za rastersku grafiku. Prvobitno je bio namijenjen za upotrebu na operativnim sustavima Windows i OS/2 za prikaz slike neovisan o uređaju. BMP format može prikazati dvodimenzionalnu

³⁸ Ibid. URL: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000018.shtml>. (21. lipnja 2018.)

³⁹ JPEG File Interchange Format. Wikipedia. URL: https://en.wikipedia.org/wiki/JPEG_File_Interchange_Format. (21. lipnja 2018.)

⁴⁰ Arms et al., Sustainability of Digital Formats, n. dj. URL: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000188.shtml>; (22. lipnja 2018.)

sliku proizvoljne visine, širine i razlučivosti kao sliku u boji dubine 1 do 64 bita ili u sivoj skali. Boje u ovom formatu mogu biti indeksirane. Dodatne mogućnosti uključuju sažimanje bez gubitaka, kanale za prozirnost i ICC profile boje. Većina BMP slika nije sažeta, a one bitne dubine 16 i 32 bita nikada nisu sažete.⁴¹ Datoteke ovog formata imaju nastavak **.bmp**.

3.2.7. GIF format

Graphics Interchange Format (GIF) je rasterski format koji je razvila tvrtka CompuServe Inc. i koji se najviše koristi za prikaz slika na Internetu. Posebno prikladnim za upotrebu na Internetu čini ga značajka da podržava tzv. progresivni prikaz, odnosno da se slika ispisuje po linijama tako da je prepoznatljivu sliku moguće vidjeti već prije nego se cijela datoteka preuzela. Ovaj format omogućuje i kratke animacije. Slika je sažeta LZW algoritmom a paleta boja sadrži najviše 256 nijansi.⁴² Datoteke ovog formata imaju nastavak **.gif**.

3.2.8. SVG format

Scalable Vector Graphics je jezik za prikaz dvodimenzionalne grafike u XML-u. SVG jezik podržava tri vrste objekata: vektorske oblike, rastersku grafiku i tekst. Ovi objekti mogu se grupirati, preoblikovati, na njih se mogu primjenjivati različiti stilovi te ih se može umetati u ranije iscrtane objekte. SVG se također može koristiti i za animacije.⁴³ Datoteke ovog formata imaju nastavak **.svg**.

3.2.9. DXF format

DXF je format za razmjenu sadržaja AutoCAD Drawing datoteka (DWG). Datoteka DXF formata je ASCII tekst organiziran kao niz skupina od kojih se svaka sastoji od dva retka. Prvi redak skupine označava kôd skupine i najčešće je pozitivan cijeli broj. Nula i mali negativni brojevi imaju posebno značenje. Drugi redak predstavlja vrijednost skupine, a format ovisi o kodu skupine u prvom retku. Skupne su organizirane u sekcije. DXF format definira sedam tipova sekcija. Prva sekcija je zaglavlje (engl. *header*) koja sadrži verziju DWG formata koja je opisana u datoteci. Osim toga ona sadrži i jedinice mjere, stilove za iscrtavanje, granice crteža, koordinatni

⁴¹ Ibid. URL: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000189.shtml>. (22. lipnja 2018.)

⁴² Ibid. URL: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000133.shtml>. (22. lipnja 2018.)

⁴³ Ibid. URL: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000020.shtml>. (23. lipnja 2018.)

sustav za prostorno pozicioniranje, naziv projekta, datume i sl. Drugi tip sekcije su razredi (engl. *classes*). On sadrži informacije o razredima čiji objekti se koriste u kasnijim sekcijama. Treći tip su tablice (engl. *tables*) koje podržavaju funkcionalnosti CAD aplikacije i koje određuju iscrtavanje sadržaja. Primjerice, u tablicama su pohranjeni stilovi crta i teksta. Četvrti tip sekcije su blokovi (engl. *blocks*). Blok je imenovan skup grafičkih objekata koji se često ponavlja na crtežu. Peti tip sekcije su grafički objekti (engl. *entities*) koji uključuju jednostavne objekta poput crte, kruga ili točke i složenije poput oblika ili zakrivljenih crta i sl. Šesti tip su objekti (engl. *objects*) poput rječnika, postavki i sl., a koriste ih CAD aplikacije za iscrtavanje. Sedmi tip sekcije je neobavezan i sadrži sličicu prikazanu uz naziv datoteke (engl. *thumbnail image*).⁴⁴ Datoteke ovog formata imaju nastavak **.dxf**.

3.2.10. EPS format

Encapsulated PostScript (EPS) format je jedan od najstarijih formata nastao krajem osamdesetih godina prošlog stoljeća. Format je razvila tvrtka Adobe System Inc. kako bi olakšala ugrađivanje ilustracija u tekstualne dokumente za ispis. Posljednja inačica specifikacija za EPS format nosi oznaku 3.0 i Adobe neće razvijati nove inačice. Format opisuje samo jednu stranicu pravokutnog oblika i za to koristi PostScript programski jezik. Posebno je prikladan za vektorsku grafiku i za kombinaciju vektorske i rasterske grafike. Zbog toga je i dalje u širokoj upotrebi u znanstvenim publikacijama. Iako format podržava i čistu rastersku grafiku, takve je primjene sve manje zbog pojave novih i prikladnijih formata.

Tipična EPS datoteka sastoji se od PostScript koda koji opisuje sliku na način da se može odaslati PostScript pisaču bez gubitaka informacije te od sličice za pregled sadržaja (engl. *preview*). Sličica za pregled sadržaja pohranjena je u formatu koji može prikazati većina grafičkih aplikacija, a obično je u nižoj razlučivosti i manjoj bitnoj dubini. Budući da sličice za pregled sadržaja mogu biti u različitim formatima, preporuča se da takve datoteke imaju različite nastavke u svom nazivu. Pokušaj otvaranja EPS datoteke u mnogim grafičkim aplikacijama otvorit će sličicu za pregled sadržaja umjesto glavnog sadržaja, a korisnik uglavnom neće biti toga svjestan.

U posljednje vrijeme otkriveni su i zloupotrijebljeni neki sigurnosni propusti u ovim datotekama pa je Microsoft u svojim Office proizvodima onemogućio uvoz ovih

⁴⁴ Ibid. URL: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000446.shtml>. (25. lipnja 2018.)

datoteka bez da se prije toga naprave određene promjene u postavkama operativnog sustava.⁴⁵ Datoteke ovog formata imaju nastavak **.eps**, **.epi**, **.epsf** ili **.epsi**.

3.2.11. SHP format

ESRI Shapefile datoteka pohranjuje netopološke strukture i informacije o prostornim parametrima objektima. Pojam *netopološki* znači da objekti pohranjeni u datoteci nemaju zajedničkih elemenata, odnosno da promjena jednog objekta koji se dodiruje sa susjednim neće uzrokovati promjenu tog susjednog objekta, odnosno da je dodirna crta pohranjena dvaput – jednom u prvom objektu, drugi puta u drugom objektu. Kod topoloških struktura dodirna crta pohranjena je samo jednom pa njena promjena mijenja oba objekta.

SHP datoteka je zapravo skup datoteka koji se sastoji od najmanje tri datoteke – glavne datoteke, indeksne datoteke i dBASE tablice. Glavna datoteka sadrži geometrijske objekte predstavljene skupom vektorskih koordinata. Svaki zapis u ovoj datoteci predstavlja oblik s listom njegovih krajnjih točaka. Indeksna datoteka sadrži informaciju o tome koliko je zapis u glavnoj datoteci udaljen od njenog početka za svaki oblik na slici. dBASE tablica sadrži podatke o obilježjima svakog oblika u glavnoj datoteci.⁴⁶ Datoteke ovog formata imaju nastavak **.shp** za glavnu datoteku, **.shx** za indeksnu datoteku i **.dbf** za dBASE tablicu.

3.2.12. GeoTIFF format

GeoTIFF format temelji se na TIFF 6.0 inačici TIFF formata i povezuje rastersku sliku s poznatim prostornim modelima ili kartografskim projekcijama. GeoTIFF format koristi se skupom TIFF oznaka za opis kartografskih podataka. Moguće je pohraniti niz georeferencijskih podataka poput tipa projekcije, koordinatnog sustava, kartografskih datuma, elipsoida i sl. Slično kao i TIFF datoteka, GeoTIFF datoteka ograničena je na veličinu od 4 GB. Veće datoteke moguće je pohraniti u BigTIFF formatu.⁴⁷ Datoteke ovog formata imaju nastavak **.tif** ili **.tiff**.

⁴⁵ Ibid. URL: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000246.shtml>. (26. lipnja 2018.)

⁴⁶ Ibid. URL: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000280.shtml>. (26. lipnja 2018.)

⁴⁷ Ibid. URL: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000279.shtml>. (26. lipnja 2018.)

3.2.13. GeoPDF format

GeoPDF je proširenje Adobe PDF formata i omogućava bilježenje georeferenciranih podataka uz sliku na stranici PDF dokumenta. GeoPDF definira područja karte povezane s PDF stranicom i preslikava mjesto na stranici u kartografski koordinatni sustav. Na taj se način za svaku točku na karti ugrađenoj u PDF stranicu izračunava geografska koordinata. Ovakav način georeferenciranja također omogućuje i izračun udaljenosti na kartama u PDF dokumentu. Budući da se radi samo o proširenju PDF formata dokumenti najčešće imaju nastavak **.pdf**.⁴⁸

3.2.14. GeoJPEG2000 format

GeoJPEG2000 format je prošireni JPEG2000 format koji omogućuje uključivanje geoprostornih metapodataka u georeferencirane slike. Ove slike mogu se čitati s bilo kojom aplikacijom koja čita JPEG2000 format, ali se može dogoditi da se geoprostorni podaci tom prilikom ne pročitaju ukoliko aplikacija ne podržava metapodatke koji sadrže ove informacije.⁴⁹ Budući da se radi samo o proširenju JPEG2000 formata dokumenti najčešće imaju nastavak **.jp2**.

⁴⁸ Ibid. URL: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000312.shtml>. (26. lipnja 2018.)

⁴⁹ GeoJP2. Wikipedia. URL: <http://fileformats.archiveteam.org/wiki/GeoJP2>. (27. lipnja 2018.)

Tablica 5: Usporedba formata za pohranu slikovnog gradiva prema tehničkim kriterijima

Format	Otkrivenost	Usvojenost	Transparentnost	Samodokumentiranje	Vanjske ovisnosti	Utjecaj patenata	Mehanizmi tehničke zaštite
TIFF	U potpunosti dokumentiran. Razvijen u Aldusu (danas u vlasništvu Adobea) i Microsoftu.	Pojedini podtipovi formata su široko rasprostranjeni kao glavni format za skenirane slike. Mnoge institucije su ga usvojile kao preporučeni arhivski format za nepokretne slike.	Ovisna je o kodiranju sadržaja, budući da je format samo omotač.	TIFF datoteka sastoji se od zaglavlja (IFH), direktorija (IFD) i samih slikovnih podataka. Svaki direktorij sadrži jednu ili više oznaka koje sadrže informacije o slici.	Nema.	Licence nisu potrebne.	Nema.
JPEG2000	Otvoreni standard koji je razvila Joint Technical Committee.	Ovaj format zauzima sve veći udio u formatima za digitalizaciju gradiva.	Ovisi o algoritmima i alatima za čitanje. Izrada alata za čitanje je kompleksna.	Manji skup metapodataka je obavezan, ali može ga se proširiti dodatnim informacijama.	Nema.	Ovise o načinu na koji je kodiran sadržaj.	Postoji polje koje donosi informacije o upravljanju intelektualnim pravima koje se može upotrijebiti za upravljanje pravom pristupa.
PNG	Otvoreni standard.	Nije široko usvojen i gubi na važnosti nakon istjecanja patentnih prava na LZW algoritam sažimanja. Ipak neke institucije preporučuju ovaj format za nepokretne rasterske slike (National Archives of Australia i Library and Archives Canada).	Ovisi o algoritmima i alatima za čitanje. Izrada alata za čitanje je kompleksna.	Dozvoljava ugradnju tekstualnih oznaka u datoteku koja sadrži nekolicinu standardnih oznaka.	Nema.	Nema.	Nema.

Format	Otkrivenost	Usvojenost	Transparentnost	Samodokumentiranje	Vanjske ovisnosti	Utjecaj patenata	Mehanizmi tehničke zaštite
JPEG/JFIF	U potpunosti otvoren.	Vrlo široko usvojen.	Transparentan.	Ograničeno, uz neke tehničke metapodatke koji su dostupni i neke koje je moguće dohvatiti na druge načine.	Nema.	Nema.	Nema.
DNG	Potpuno dokumentiran. Razvio ga je Adobe Systems, Inc.	Format nije široko usvojen. Usvojili su ga tek neki proizvođači digitalnih kamera, ali oni najveći nisu.	Omotač je transparentan, a za kodirane slike koje sadrži može biti potreban alat za prikaz.	Metapodaci mogu biti ugrađeni kao TIFF ili EXIF.	Nema.	Adobe dozvoljava upotrebu specifikacije za razvoj i promicanje DNG formata.	Nema.
BMP	Zaštićeni format.	Primarno je zamišljen za upotrebu na Windows operativnim sustavima, ali se zbog jednostavne strukture i nesažetosti proširio pa ga podržavaju svi Internet preglednici.	Jednostavan rasterski format. Iako je sažimanje podržano, slike u ovom formatu nisu sažete.	Nema.	Nema.	Nema.	Nema.
GIF	Zaštićeni format, ali javno dostupne specifikacije.	Široko usvojen.	Relativno transparentan, ali transparentnost ovisi o algoritmima za kodiranje.	Nema.	Nema.	Nema.	Nema.
SVG	Otvoreni standard.	Podržavaju ga Adobeovi alati, Microsoft Visio, Corel Draw 11 i drugi. Podržavaju ga gotovo svi mrežni preglednici.	Transparentan, budući da se temelji na XML jeziku.	Može sadržavati bogati skup metapodataka budući da koristi elemente iz vanjskih XML shema.	Nema.	Nema.	Nema.

Format	Otkrivenost	Usvojenost	Transparentnost	Samodokumentiranje	Vanjske ovisnosti	Utjecaj patenata	Mehanizmi tehničke zaštite
DXF	Zaštićeni format, ali javno dostupne specifikacije.	U vrlo širokoj upotrebi za razmjenu podataka između CAD aplikacija. Međutim njegova upotreba danas je sve manja s obzirom da današnje CAD aplikacije imaju puno više funkcionalnosti od jednostavne dvodimenzionalne grafike.	Jednostavan tekst format može se otvoriti s bilo kojim osnovnim alatom za pregled teksta. Ipak, mnogi elementi su označeni brojčanim kodovima koje je teško interpretirati.	Nema.	Nema.	Nema.	Nema.
EPS	Zaštićeni format, ali javno dostupne specifikacije.	EPS je jedan od najstarijih vektorskih formata iz kasnih 1980-ih godina. Od tada je bio u širokoj upotrebi, međutim danas ga zbog nedostataka potiskuju noviji formati.	U suštini EPS je tekstualna datoteka koja koristi PostScript programski jezik. Čitljiva je ljudima, ali je često potreban alat zbog kompleksno zapisane grafike.	Sadrži neke unaprijed definirane metapodatke, ali moguće ga je povezati s elementima vanjskih XML shema.	Nema.	Nema.	Nema.
SHP	U cijelosti dokumentiran. Specifikacija je javna.	Danas je u širokoj upotrebi, unatoč nekim manjkavostima. Mnoge institucije svoje podatke pohranjuju u ovom formatu (U.S. Geological Survey, U.S. Census Bureau, Environmental Protection Agency i dr.)	Moguće je razviti računalne programe za čitanje ili zapis u SHP datoteke.	GIS metapodaci mogu se uključiti u .xml datoteku u okviru datoteka koje čine ovaj format.	Nije poznato.	ESRI Licence Agreements.	Nije poznato.

Format	Otkrivenost	Usvojenost	Transparentnost	Samodokumentiranje	Vanjske ovisnosti	Utjecaj patenata	Mehanizmi tehničke zaštite
GeoTIFF	Otvoren i u cijelosti dokumentiran.	Široko rasprostranjen. Upotrebljavaju ga brojne institucije.	Ovisna je o kodiranju sadržaja, budući da je format samo omotač.	Uz sve što je navedeno za TIFF format, GeoTIFF pohranjuje metapodatke u nekoliko dodatnih TIFF oznaka.	Za pregled je potrebno koristiti aplikaciju koja može interpretirati metapodatke iz dodatnih TIFF oznaka.	Dijelovi specifikacije zaštićeni su intelektualnim pravima, ali ih je dopušteno koristiti u nekomercijalne svrhe uz navođenje autorskih prava.	Nije poznato.
GeoPDF	Open Geospatial Consortium je prihvatio i objavio specifikacije.	Brojne američke institucije upotrebljavaju ovaj format.	Geografske metapodatke moguće je čitati s osnovnim tekstualnim alatima.	Sam GeoPDF u PDF dokument unosi ograničen skup geoprostornih metapodataka.	Potrebno je instalirati dodatke (engl. <i>add-on</i>) u aplikacije za čitanje PDF dokumenata kako bi se u potpunosti mogle iskoristiti funkcionalnosti ovog formata.	Sam pojam GeoPDF može se upotrijebiti samo na proizvodima koji su napravljeni po Terrago licenci. Međutim specifikacija je objavljena kao besplatna kako bi drugi mogli razvijati aplikacije koje se koriste ovim sustavom bilježenja geografskih metapodataka.	Nema. Enkripcija se može primijeniti na PDF dokument koji su sebi sadrži GeoPDF metapodatke.
GeoJPEG2000	Specifikacije su javno dostupne.	Nema podataka.	Budući da je format utemeljen na GeoTIFF formatu vrijede ista pravila.	Budući da je format utemeljen na GeoTIFF formatu vrijede ista pravila.	Nema podataka.	Nema. Preporuka je da se ne koristi naziv GeoJP2 zbog zaštite žiga.	Nema.

Tablica 6: Usporedba preporučenih formata za pohranu nepokretnih slika grade prema kriterijima kvalitete i funkcionalnosti

Format	Ispravan prikaz	Jasnoća	Očuvanje boje	Podrška za grafičke efekte i tipografiju	Podrška za višespektralne pojaseve
TIFF	Dobro podržava.	Izvršno podržava slike vrlo visoke razlučivosti. Prostor boja i bitna dubina su fleksibilni.	Oznaka za ICC prostor boja uređaja koji generira sliku dodana je u TIFF/IT i TIFF/EP inačice. Proširene oznake mogu se koristiti u svim inačicama TIFF obitelji. Nema jamstva da će sve aplikacije za čitanje TIFF formata pročitati skup proširenih oznaka.	Ne podržava vektorsku grafiku.	Nema.
JPEG2000	Dobro podržava.	Izvršno podržava slike velikog formata i visoke razlučivosti i bitne dubine. Komponente boje mogu se transformirati bez ili uz gubitke, a podržan je širok spektar prostora boja. Sažimanje može biti bez ili uz gubitke.	Bogata podrška. Prostor boja naznačen je u posebnom polju ako se koristi, ili se koristi paleta boja. Boju je moguće razdvojiti na komponente uključujući i komponentu prozirnosti.	Ne podržava vektorsku grafiku.	Nema.
PNG	Dobro podržava.	Izvršno podržava uz podršku progresivnom prikazu slika dok se učitavaju. Prostor boja i bitna dubina su fleksibilni, a boje mogu biti i indeksirane.	Datoteka može sadržavati informacije za gama korekciju.	Podatak o prozirnosti može se uključiti na razini svake pojedine točke za slike u boji ili u sivoj skali.	Nema.
JPEG/JFIF	Dobro podržava.	Dobra, ali podaci su ograničeni na 8 bita po kanalu boje. Kvaliteta ovisi o složenosti algoritma za kodiranje i stupnju sažimanja.	Ograničena podrška za prostor boja.	Ne podržava vektorsku grafiku.	Nema.
DNG	Dobro podržava.	Visoka razina jasnoće u skladu s tehničkim mogućnostima kamere.	Izvršno, uz podršku za ICC profile boje i identifikaciju prostora boja.	Nije relevantno budući da je format orijentiran na kameru.	Nepoznato.
BMP	Podržava uvećanje i kretanje po slici.	Koristi se samo za male slike u kojima se ne zahtijeva velika bitna dubina.	Podržava ICC profile boje.	Ne podržava vektorsku grafiku ni tipografiju.	Nema.

Format	Ispravan prikaz	Jasnoća	Očuvanje boje	Podrška za grafičke efekte i tipografiju	Podrška za višespektralne pojaseve
GIF	Dobro podržava.	Dobra, u granicama paletnih boja.	Izvršno, budući da GIF format sadrži informacije o paleti boja u obliku RGB vrijednosti.	Nema.	Nema.
SVG	Dobro podržava.	Moguće je uvećavati i smanjivati sliku bez gubitka jasnoće.	Moguće je specificirati ICC profile boje.	Dobro podržava.	Nije primjenjivo.
DXF	Dobro podržava.	Moguće je uvećavati i smanjivati sliku bez gubitka jasnoće.	Prostor boja je indeksiran ili u RGB prostoru dubine 24 bita.	Izvršno podržava vektorsku grafiku. Definirano je više od 60 vrsta geometrijskih oblika koje je moguće grupirati ili smjestiti u slojeve.	Nije primjenjivo.
EPS	Dobra podrška za kretanje po slici, uvećavanje i ispis. Nema podrške za slojeve i prozirnost.	Moguće je uvećavati i smanjivati sliku bez gubitka jasnoće.	Dobra podrška za definiciju prostora boja kao i mijenjanje prostora boja dok se slika iscrtava.	Odlično podržava vektorsku grafiku.	Nije primjenjivo.
SHP	Nije primjenjivo.	Nije primjenjivo.	Nije primjenjivo.	Nije primjenjivo.	Nije primjenjivo.
GeoTIFF	Nije primjenjivo.	Nije primjenjivo.	Nije primjenjivo.	Nije primjenjivo.	Nije primjenjivo.
GeoPDF	Nije primjenjivo.	Nije primjenjivo.	Nije primjenjivo.	Nije primjenjivo.	Nije primjenjivo.
GeoJPEG2000	Nije primjenjivo.	Nije primjenjivo.	Nije primjenjivo.	Nije primjenjivo.	Nije primjenjivo.

3.3. Formati za zvukovnu građu

Kada promatramo formate koji sadrže zvukovni sadržaj možemo ih podijeliti u dvije skupine. Prva skupina odnosi se na formate koji predstavljaju snimljeni zvuk, odnosno valni oblik zvučnog signala. Druga skupina obuhvaća formate namijenjene generiranju zvuka kroz softver ili hardver, ili njihovu kombinaciju. Neki od formata iz ove druge skupine obrađeni su u poglavlju o tekstualnim formatima za notno gradivo. Ti formati osim za pripremu ispisa nota mogu poslužiti i za razmjenu podataka s aplikacijama i uređajima za sintezu zvuka. Ovo poglavlje usredotočit će se na prvu skupinu formata za zvukovni sadržaj.

Prilikom odabira najpovoljnijeg formata za zvukovnu građu treba se voditi sljedećim kriterijima kvalitete i funkcionalnosti:

Ispravna reprodukcija zvuka odnosi se na reprodukciju zvuka u mono ili stereo tehnici kroz jedan ili dva zvučnika ili, u slučaju *surround* zvuka kroz nekoliko zvučnika raspoređenih po prostoriji. Ona također obuhvaća programsku kontrolu glasnoće, boje tona, balansa i drugih parametara zvuka kao i mogućnost brzog preslušavanja i traženja određenih mjesta u snimljenom materijalu. Uobičajena reprodukcija uključuje i mogućnost analize zvuka i uzimanje isječaka iz cjeline. Ova funkcionalnost ne smije biti ograničena korištenom tehnologijom odnosno modelom uređaja te mora biti dostupna svim korisnicima.

Vjernost reprodukcije veže se uz stupanj do kojeg se HiFi (engl. *High Fidelity*) sadržaj može reproducirati nekim formatom. Ovo je više subjektivan kriterij kada govorimo o reprodukciji, budući da tek osjetljivo i uvježbano uho može primijetiti nijanse u vjernosti reprodukcije. Međutim, on postaje mnogo važniji ukoliko pohranjeni materijal postaje temelj za novo izdanje zvukovnog medija (CD, gramofonska ploča), kada će kvaliteta novog izdanja uvelike ovisiti o tome koliko je vjerno izvorni materijal prenesen u digitalni format.

Dvije značajke koje se najčvršće vežu uz vjernost zvučnog signala predstavljenog linearnom impulsno kodnom modulacijom (LPCM) su frekvencija uzorkovanja i bitna dubina. Uz to na vjernost mogu utjecati i drugi čimbenici poput izobličenja, dodavanja vodenih žigova zvučnom zapisu i sažimanja uz gubitke. U pravilu najvjernije rezultate daje zapis koji nije sažet ili je sažet bez gubitaka, te onaj koji je uzorkovan višom frekvencijom i s većom bitnom dubinom. Ipak, zapisi koji su

sažeti s gubitkom koji u obzir uzima karakteristiku ljudskog doživljaja zvuka još uvijek mogu dostići zadovoljavajuću razinu vjernosti u normalnim uvjetima reprodukcije.

Višekanalna podrška odnosi se na to koliko uspješno formati mogu predstavljati višekanalni zvuk, koji se slušatelju može prenijeti u barem dva oblika. Prvi je u obliku zvučnog prostora, odnosno zvučnog polja kao stereo ili *surround* zvuk. Drugi oblik je reprodukcija zvuka iz dva izvora u cilju osiguranja alternativnih ili dodatnih sadržaja, na primjer čitanje teksta na dva različita jezika ili zvučni efekti koji nisu sastavni dio glazbe ili primjerice tzv. Music Minus One (MMO) – zvučni zapis iz kojega je uklonjena jedna dionica – glas ili instrument, kako bi tu dionicu slušatelj mogao sam nadomjestiti u svrhu vježbe ili zabave.

Podrška za preuzete ili korisnički definirane zvukove pokazuje u kojoj mjeri format dozvoljava uključivanje ili povezivanje s digitalnim zvučnim podacima potrebnim za stvaranje jednog ili više glasova ili instrumenata u glazbenoj prezentaciji.⁵⁰

Za zvučno gradivo Kongresna knjižnica preporučuje formate i metapodatke prikazane u tablici.⁵¹

⁵⁰ Arms et al., Sustainability of Digital Formats, n. dj. URL: http://www.loc.gov/preservation/digital/formats/content/sound_quality.shtml. (12. lipnja 2018.)

⁵¹ Library of Congress Recommended Formats Statement, n. dj.

Tablica 7: Preporučeni formati za zvukovno gradivo

Vrsta zvukovnog gradiva	Preporučeni format	Metapodaci
Digitalni zvučni zapisi neovisni o mediju	PCM WAVE datoteka najviše razlučivosti (44,1 kHz /16 bit ili bolja) s ugrađenim metapodacima (Broadcast WAVE). Datoteke vrlo visoke razlučivosti poput DSD, PCM 176,4 kHz, 192 kHz do 384 kHz ako su proizvedene za objavu uz CD ako postoje oba izdanja. DSD objavljenog zapisa. MP3 uzorkovan s 128 kbit/s (mono), odnosno 256 kbit/s (stereo) ili više.	Naziv pjesme/djela Naziv albuma Izvođač Skladatelj Glazbeni žanr Izdavač/Naziv izdavača i broj izdanja Mjesto i datum izvođenja Datum izdanja Standardni identifikator (npr. ISRC, UPC) Ostali identifikatori Podaci trebaju biti u standardnom formatu utemeljenom na XML strukturi poput ERN-DDEX formata.
Protočni (engl. <i>streaming</i>) zvuk	Nesažeti format Zvuk sažet nekom standardnom široko primijenjenom metodom sažimanja.	Nema.
Prateće slike i tekstualne datoteke	Za prateće slike uvijek se preporuča viša razlučivost. Slike u TIFF ili JPG formatu, tekst u PDF formatu.	Nema.

3.3.1. WAVE format

WAVE format koristi se za pohranu zvukovnog sadržaja na PC računalima. Razvili su ga Microsoft i IBM kao implementaciju metode RIFF formata za pohranu podataka u cjelinama (*chunks*). Protok bitova obično je kodiran linearnom impulsno kodnom modulacijom (LPCM). Iako je prvenstveno razvijen kao format za PC računala, WAVE format kompatibilan je s Linux i Apple računalima. Unatoč tome što ovaj format podržava i sažimanje zapisa, najčešće WAVE zapis nije sažet što ga čini

prikladnim kandidatom za pohranu digitalnog zapisa visoke kvalitete.⁵² Datoteke ovog formata imaju nastavak **.wav**.

3.3.2. Broadcast WAVE format

Ovaj format temelji se na WAVE formatu, a namijenjen je razmjeni zvukovnog materijala između različitih okruženja za difuziju signala i opreme koja radi na različitim platformama. U odnosu na WAVE format ovom formatu dodana je cjelina (*chunk*) koja sadrži osnovne metapodatke koji se smatraju nužnima za primjenu u difuziji signala. Uz ove osnovne metapodatke, razvijene su i dodatne cjeline koje sadrže dodatne metapodatke.⁵³ Datoteke ovog formata imaju nastavak **.wav**.

3.3.3. DSD format

DSD (*One-bit Delta Sigma*, odnosno trgovački naziv *Direct Stream Digital*) modulacija zvuka je način kodiranja zvuka modulacijom gustoće impulsa. Signal se pohranjuje kao niz vrijednosti dubine jednog bita i visoke frekvencije. Iako upotreba ovog formata nije ograničena samo na zvukovnu građu, u ovom odjeljku obradit će se samo taj aspekt primjene. Frekvencija uzorkovanja 64 puta je viša od frekvencije uzorkovanja koja se koristi kod audio CD medija, odnosno 2,8224 MHz. Međutim, bitna dubina je samo jedan bit. Zagovornici ovog formata tvrde da prilikom obrade ovako uzorkovanog signala dolazi do puno manje izobličenja i šuma nego kod signala nastalog primjenom LPCM modulacije.⁵⁴

3.3.4. MP3 format

MP3 uobičajen je naziv za treći sloj kodiranja zvuka u MPEG formatu. (engl. *MPEG Layer III audio encoding*). MP3 format sažima zvučni zapis tzv. perceptualnim kodiranjem, odnosno primjenom psihoakustičkog modela koji uklanja ili stišava elemente sadržaja koje ljudsko uho ne čuje ili čuje slabije. MPEG format može kodirati zvuk na tri razine, odnosno tri sloja. Svaki viši sloj donosi veće sažimanje zvuka, ali uz veće smanjenje njegove kvalitete. MP3 datoteka kodirana s 128 kbit/s zauzima oko 1/11 veličine nesažetog WAV formata CD kvalitete.⁵⁵ MP3 datoteke imaju nastavak **.mp3**.

⁵² Arms et al., Sustainability of Digital Formats, n. dj. URL:

<http://www.loc.gov/preservation/digital/formats/fdd/fdd000002.shtml>. (13. lipnja 2018.)

⁵³ Ibid. URL: <http://www.loc.gov/preservation/digital/formats/fdd/fdd000357.shtml>. (13. lipnja 2018.)

⁵⁴ Ibid. URL: <http://www.loc.gov/preservation/digital/formats/fdd/fdd000230.shtml>. (13. lipnja 2018.)

⁵⁵ Ibid. URL: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000012.shtml>. (05. srpnja 2018.)

Tablica 8: Usporedba formata za pohranu zvukovnog gradiva prema tehničkim kriterijima

Format	Otkrivenost	Usvojenost	Transparentnost	Samodokumentiranje	Vanjske ovisnosti	Utjecaj патената	Mehanizmi tehničke zaštite
WAVE	U potpunosti dokumentiran. Razvijen u Microsoftu i IBM-a kao dio RIFF formata, a dokumentacija je javno dostupna.	Široko usvojen format. Uz LPCM kodiranje ovaj se format preporučuje za dugoročno očuvanje zvukovnog gradiva u mnogim arhivima.	Ovisi o primijenjenom načinu kodiranja. U slučaju LPCM kodiranja transparentnost je visoka te se signal lako može dekodirati. ⁵⁶	Metapodaci se mogu smjestiti u INFO cjelinu (<i>chunk</i>)	Nema.	Licence nisu potrebne.	Nema.
Broadcast WAVE	Otvoreni standard koji je razvila EBU (<i>European Broadcast Union</i>) na temelju WAVE formata.	Široko preporučen. Preporučeni format za isporuku glazbenog materijala izdavačkim kućama.	Ovisi o primijenjenom načinu kodiranja. U slučaju LPCM kodiranja transparentnost je visoka te se signal lako može dekodirati. Metapodaci su čitljivi čovjeku.	Broadcast Audio Extension cjelina sadrži metapodatke za identifikaciju podrijetla i datuma i vremena nastanka, kao i tehničko administrativne metapodatke.	Nema.	Nema.	Nema.
DSD	Koncept je široko obrađen literaturi. Format je razvio SONY.	Stupanj usvojenosti je nepoznat. SONY ga upotrebljava u svojim SACD (<i>Super Audio CD</i>) diskovima, ali ne postoji široka prihvaćenost formata na glazbenom tržištu.	Ovaj format predstavlja nesažeti prikaz zvukovnog materijala te je kao takav relativno transparentan.	Nema.	Nema.	Licenciranju podliježe SACD. Konceptualni model ne podliježe licenciranju.	Nema.

⁵⁶ Ibid. URL: <http://www.loc.gov/preservation/digital/formats/fdd/fdd000011.shtml>. (13. lipnja 2018.)

Format	Otkrivenost	Usvojenost	Transparentnost	Samodokumentiranje	Vanjske ovisnosti	Utjecaj patenata	Mehanizmi tehničke zaštite
MP3	Otvoren standard.	U širokoj upotrebi za razmjenu sadržaja putem Interneta i reprodukciju na posebnim uređajima.	Za čitanje je nužna upotreba algoritama i alata.	Tehničke informacije (o kodiranju) nalaze se u zaglavlju „okvira“ koji tvore tok bitova. Formatu nedostaju opisni metapodaci pa se često koristi posebno razvijena ID3 struktura metapodataka.	Nema.	Patenti vezani uz MP3 format istekli su u travnju 2017.	Nema.

Tablica 9: Usporedba preporučenih formata za pohranu zvukovne građe prema kriterijima kvalitete i funkcionalnosti

Format	Ispravna reprodukcija zvuka	Vjernost reprodukcije	Višekanalna podrška	Podrška za preuzete ili korisnički definirane zvukove
WAVE	Izvršno podržava.	Ovisi o načinu kodiranja. Uz kodiranje LPCM modulacijom dobiva se nesažeti signal. Vjernost se podiže povećanjem frekvencije uzorkovanja i bitne dubine.	Višekanalnost se može postići i upotrebom WAVE Format Extensible formata.	Nema.
Broadcast WAVE	Podržava.	Ovisi o načinu kodiranja. Uz kodiranje LPCM modulacijom dobiva se nesažeti signal. Vjernost se podiže povećanjem frekvencije uzorkovanja i bitne dubine.	Ovisi o načinu kodiranja. LPCM kodira samo jedan kanal zvuka. Stereo zvuk postiže se preplitanjem uzoraka drugog kanala između uzoraka prvog. Višekanalnost se može postići i upotrebom WAVE Format Extensible formata.	Nema.
DSD	Dobro podržava.	Format podržava visoku razlučivost zvuka. Primijenjen na SACD s frekvencijom uzorkovanja od 2,8 MHz omogućeno je bilježenje frekvencija do 100 kHz uz dinamički raspon od 120 dB.	Višekanalnost nije temeljno obilježje toka bitova (<i>bitstream</i>), ali se datoteke mogu strukturirati tako da predstavljaju 2, 5 ili 6 kanala.	Nema.
MP3	Dobro podržava.	Osrednja do dobra ako se uzme u obzir da se radi o sažetom formatu. Kvaliteta ovisi o vrsti i stupnju sažimanja te o softveru za kodiranje.	Podržava pet osnovnih kanala i jedan za niskotonske efekte.	Nema.

4. Zaključak

Uz brz razvoj informatičke tehnologije krajem 20. stoljeća, podjednako brzinom se razvijaju i različiti načini pohrane podataka. To je dovelo do toga da danas postoji ogroman broj različitih formata za pohranu informacija. Velik broj tih formata više se ne koristi, budući da su ih zamijenili napredniji, međutim neki se i dalje koriste i razvijaju.

Budući da je nemoguće u okviru diplomskog rada opisati sve danas korištene formate za pohranu različitih sadržaja, morao sam se ograničiti na one koji čuvaju one vrste sadržaja koje najčešće nalazimo u knjižnicama i arhivima. Zbog toga sam se usredotočio na samo tri vrste sadržaja – tekst (popraćen slikama ili bez njih), nepokretne slike i zvuk. Međutim, čak i uz ovakvo ograničenje nemoguće je opisati sve vrste formata koje za ove sadržaje danas postoje.

Općenito, kako bi se neki format smatralo prikladnim za dugoročno očuvanje gradiva on mora zadovoljiti što je više moguće od sedam općih tehničkih zahtjeva te dodatne zahtjeve koje određuje sama vrsta sadržaja.

Kada promatramo tekstualno gradivo i njegovo dugoročno očuvanje u digitalnom obliku moramo odlučiti želimo li dokument sačuvati izgledom, dakle želimo li da svaka stranica izgleda točno onako kako izgleda u izvorniku, ili nam je bitan samo sadržaj bez obzira na njegovu prezentaciju. O tom izboru ovisit će i odabir vrste formata.

Za formate koji ne čuvaju izgled stranice usporedio sam formate BITS i EPUB. Iako oba formata udovoljavaju svim tehničkim uvjetima za dugoročnost očuvanja gradiva, prednost je na BITS formatu. To je zbog toga što u slučaju EPUB formata dodatno treba voditi računa da se ne koristi šifriranje sadržaja, što format sam po sebi dozvoljava, čime bi se narušio uvjet transparentnosti. Drugi nedostatak EPUB formata je u uvjetu samodokumentiranja odnosno količine i prikladnosti oznaka koje sadrže metapodatke. BITS format ima bogatu paletu oznaka koje su sastavni dio formata, dok je EPUB nešto skromniji te za veći izbor metapodataka treba dokument povezati s metapodacima pohranjenim izvan samog formata. Osim toga, EPUB omogućava upravljanje intelektualnim pravima, pa kod pripreme sadržaja treba voditi računa da se ova mogućnost ne uključi. Još jedan nedostatak EPUB formata je njegova potencijalna nepostojanost u smislu specifikacija. Oba su formata relativno nova, ali je BITS format postojaniji, njegova druga inačica ne razlikuje se puno od prve i prošlo je oko četiri

godine između pojave dviju inačica. Za razliku od toga treća inačica EPUB formata u potpunosti se razlikuje od druge, a treća se inačica pojavila samo godinu dana nakon druge. Zbog svega navedenog, BITS format nameće se kao najbolji izbor formata za dugoročno očuvanje teksta u kojemu nije bitan izgled stranice. Kao takvoga ga je prepoznala Kongresna knjižnica.

Među formatima koji čuvaju izgled stranice obitelj PDF formata nameće se svojom rasprostranjenošću. Nisu, međutim svi članovi ove obitelji jednako prikladni za pohranu dokumenata koje treba dugoročno čuvati. To je ponajprije zbog toga što PDF format u svojoj osnovnoj inačici pruža razne mogućnosti kojima izlazi iz okvira uvjeta koje format za dugoročno očuvanje dokumenata mora zadovoljiti. Tako na primjer osnovni PDF format omogućava dodavanje privitaka u obliku datoteka glavnom dokumentu čime bi se narušilo načelo vanjske neovisnosti, budući da svaki takav privitak može trebati neku specifičnu aplikaciju za čitanje, a pitanje je hoće li takva aplikacija biti na raspolaganju nakon proteka duljeg perioda. Zatim, osnovni PDF format omogućuje ugradnju i izvršavanje JavaScript koda zbog čega također može stvoriti vanjsku ovisnost. Uz to, osnovni format omogućuje šifriranje sadržaja što također nije poželjno ako želimo dugoročno čuvati dokumente. Još jedna važna osobina PDF formata je da ukoliko se želi zadržati vjerni prikaz stranice sva korištena pisma – fontovi trebaju biti ugrađeni u dokument. Zbog svih ovih potencijalnih zamki pri korištenju PDF formata u svrhu dugoročnog očuvanja dokumenata razvijene su inačice PDF formata koje ograničavaju navedene funkcionalnosti. Najvažnija inačica, koje je i razvijena baš za arhiviranje dokumenata je inačica PDF/A. Ova inačica ne dozvoljava privitke, JavaScript kôd i šifriranje dokumenta. Kao i u osnovnoj inačici, pisma moraju biti ugrađena u dokument. Prostor boja mora biti definiran tako da bude neovisan o platformi na kojoj se dokument čita. Dozvoljena je upotreba isključivo standardnih metapodataka. Osim inačice PDF/A za dugoročno očuvanje dokumenata može se upotrijebiti i format PDF/UA. Ova inačica dozvoljava šifriranje dokumenta i umetanje zvučnih elemenata pa pri njenom korištenju treba voditi računa da se ove funkcionalnosti ne koriste. Osim toga ova inačica zabranjuje da se informacija prenosi putem boje, kontrasta ili zvuka (zvuk ionako ne dolazi u obzir ako je svrha dugoročno očuvanje), a ukoliko se informacija u dokumentu prenosi putem (zvuka,) boje ili kontrasta potrebno je takve dijelove sadržaja označiti i opisati smisao slike (ili zvuka). Sve navedeno za PDF formate ističe format PDF/A kao najprikladniji za dugoročnu pohranu dokumenata. Ukoliko se povede računa da se sadržaj formata PDF/UA uskladi

sa specifikacijama za PDF/A tada i ovaj format postaje prikladan za dugoročno očuvanje. Kongresna knjižnica ovako usklađeni PDF/UA format ističe na prvom mjestu prikladnih formata zbog toga što je prilagođen upotrebi osobama s posebnim potrebama.

U skupinu formata za tekstualnu građu ubrajaju se i formati za notne zapise. Ako je potrebno sačuvati izgled stranice notnog zapisa, tada je očiti izbor neki od PDF formata prikladnih za dugoročno očuvanje dokumenata, međutim ako se želi zabilježiti samo sadržaj nekog glazbenog djela tada se ističu dva formata – MusicXML i MEI. Ova dva formata su vrlo slična sadržajem iako je MusicXML bolje strukturiran od MEI formata. To ga čini lakšim za razumijevanje i razvoj aplikacija koje ga čitaju. S druge strane MEI zapis istoga glazbenog djela je kraći. MusicXML može opisati note zapisane u zapadnom notnom sustavu iz 17. stoljeća do danas (menzuralna notacija), dok MEI format može opisati sustave koji datiraju od oko 8. stoljeća (neumatska notacija). Unatoč manjoj veličini MEI zapisa i mogućnosti zapisa starijih notnih sustava, strukturiranost i veća razumljivost MusicXML zapisa daje mu prednost pri odabiru formata.

Kada je riječ o formatima za slikovno gradivo i ovdje možemo formate podijeliti u dvije skupine. Jedna su formati za pohranu rasterskih slika, a drugi su namijenjeni pohrani vektorske grafike. Kao podskupinu svake od ovih skupina možemo navesti i formate za geoprostorne podatke, iako se zapravo radi samo o proširenju osnovnih formata dodatnim skupom metapodataka.

Kod rasterskih formata slika se sastoji od polja točaka, a svaka točka opisana je s određenim brojem bitova. Od rasterskih formata svojim značajkama za dugoročno očuvanje dokumenata može se izdvojiti nekoliko formata. Svi formati koje sam obuhvatio ovim diplomskim radom potpuno su dokumentirani i specifikacije su javno dostupne, bez obzira radi li se o otvorenim specifikacijama ili onima koje su zaštićene intelektualnim pravima. Promatramo li formate kroz kriterije samodokumentiranja, potrebno je istaknuti TIFF format i DNG format, koji za metapodatke koristi isti skup oznaka kao i TIFF. TIFF format je uz to široko rasprostranjen, a datoteka može sadržavati i više od jedne slike što mu daje dodatnu prednost. Od ostalih formata valja istaknuti PNG format koji podržava progresivno iscertavanje slika, odnosno za iscertavanje za vrijeme njihovog preuzimanja. U novije vrijeme sve je više prisutan format JPEG2000 koji po svojim obilježjima ne zaostaje za TIFF i PNG formatima. Uspoređujući sve kriterije koje formati za dugoročno očuvanje slikovnog gradiva

trebaju zadovoljavati jako je teško izdvojiti jedan format koji bi se moglo nazvati preporučenim. Međutim, ako uzmemo u obzir dosadašnju praksu, većina institucija, uključujući i Kongresnu knjižnicu još uvijek prednost daje formatu TIFF.

Vektorski formati svoj sadržaj opisuju matematički. Sva četiri promatrana formata imaju dostupnu specifikaciju i podjednako su usvojeni na tržištu. ESRI licenčni ugovori koji propisuju načine upotrebe SHP formata stavlja ovaj format u nepovoljniji položaj kad je u pitanu dugoročno čuvanje dokumentacije, dok nedostatak adekvatnih oznaka za metapodatke ne ide u prilog DXF formatu. Ako uzmemo u obzir činjenicu da je EPS format možda najstariji format za vektorsku grafiku, unatoč velikoj rasprostranjenosti zbog starosti i sigurnosnih rizika on gubi na važnosti. Sve to u prvi plan gura SVG format. To je format otvorene specifikacije, dakle neovisan o intelektualnim pravima. Temelji se na XML jeziku i samim time je transparentan, a informacije o sadržaju datoteke moguće je opisati metapodacima bilo koje poznate XML sheme.

Zvukovno gradivo za dugoročno očuvanje najčešće se priprema digitalizacijom već postojećih medija u visokoj kvaliteti. U razmatranje najpogodnijeg formata za dugoročno očuvanje zvukovnog gradiva uzeo sam četiri formata datoteka. Format WAV izrazito je raširen i potpuno je dokumentiran te slobodan za korištenje u pogledu intelektualnih prava. Njegov nedostatak je manjak informacija o zapisu koji se nalazi zapisan u datoteci. Osim toga ako se za kodiranje sadržaja koristi linearna pulsno kodna modulacija (LPCM) tada je on i transparentan. Zbog nedostatka metapodataka Europska radiodifuzijska unija (EBU) razvila je format utemeljen na WAV formatu uz dodatak bogatog skupa metapodataka za opis pohranjenog zvukovnog gradiva – Broadcast WAVE format. Ovaj format je otvoren i u potpunosti dokumentiran te nije ograničen nikakvim intelektualnim pravima. Uvjet raširenosti zadovoljava i format MP3 koji, međutim, nema pokazatelja o dostupnosti specifikacija. Također je nejasno u kojoj mjeri je algoritam kodiranja opterećen patentnim pravima. Skup metapodataka je također skroman i ograničen na tehničke podatke. Ovi nedostaci svakako eliminiraju MP3 format iz kandidature za izbor najpovoljnijeg formata za dugoročno očuvanje zvučnih zapisa. Posljednji u ovoj analizi je DSD format zbog zanimljivog načina kodiranja sadržaja dubinom uzorka od samo jednog bita dok se zvučnim signalom modulira gustoća samih uzoraka. Ovaj sustav u praktičnoj je primjeni koristi jedino Sony u svojim Super Audio CD (SACD) medijima. Koncept DSD formata nije opterećen intelektualnim pravima, i kao takav je otkriven. Količina metapodataka koji

prate ovu vrstu zapisa je ograničena. Na temelju razmatranja formata za zvukovno gradivo može se zaključiti da je preporučeni format Broadcast WAVE.

Ovim diplomskim radom obuhvatio sam tri osnovne vrste gradiva. Budući da je težište rada i sam predmet studija bibliotekarstvo, a ne arhivistika, izostavio sam pokretne slike kao vrstu građe, budući da je udio takve građe u knjižnicama, za razliku od arhiva, zanemariv.

Kroz ovaj rad opisao sam samo neke od formata za pohranu gradiva, i to one za koje sam smatrao da prema zadanim kriterijima mogu poslužiti svrsi dugoročnog očuvanja. Moramo biti svjesni da se već za nekoliko godina mogu pojaviti novi formati, kao što je to primjerice WPUB format koji bi mogao zamijeniti EPUB.⁵⁷ koji će se bolje uklopiti u zadane kriterije, a također možemo pretpostaviti da će se s razvojem tehnologije mijenjati i sami kriteriji. Potrebno je stalno pratiti razvoj formata i tehnologija te analizirati kako se novi formati i nove inačice poznatih formata uklapaju u tražene kriterije. Naravno, i same kriterije treba provjeravati i usklađivati s napretkom tehnologije, iako sam mišljenja da ta promjena ne treba biti toliko brza kao što se brzo javljaju novi formati i tehnologije. Barem za sada!

⁵⁷ Web Publications. W3C Editor's Draft 23 August 2018, <https://w3c.github.io/wpub/>. (3. rujna 2018.)

5. Popis tablica

Tablica 1: Preporučeni formati za tekstualno gradivo.....	14
Tablica 2: Usporedba formata za pohranu tekstualne građe prema tehničkim kriterijima	22
Tablica 3: Usporedba preporučenih formata za pohranu tekstualne građe prema kriterijima kvalitete i funkcionalnosti	27
Tablica 4: Preporučeni formati za nepokretne slike	30
Tablica 5: Usporedba formata za pohranu slikovnog gradiva prema tehničkim kriterijima	38
Tablica 6: Usporedba preporučenih formata za pohranu nepokretnih slika građe prema kriterijima kvalitete i funkcionalnosti	42
Tablica 7: Preporučeni formati za zvukovno gradivo	46
Tablica 8: Usporedba formata za pohranu zvukovnog gradiva prema tehničkim kriterijima	48
Tablica 9: Usporedba preporučenih formata za pohranu zvukovne građe prema kriterijima kvalitete i funkcionalnosti	50

6. Literatura

1. A music notation engraving library. URL: <http://www.verovio.org/index.xhtml>. (3. lipnja 2018.)
2. ahankinson. Music Encoding Initiative. 5. veljače 2015. <https://musescore.org/en/node/24208>. (3. lipnja 2018.)
3. An introduction to MEI. Music Encoding Initiative. URL: <http://music-encoding.org/about/>. (3. lipnja 2018.)
4. Andreis, J. Povijest glazbe. 1975. Zagreb : Liber Mladost.
5. Arms C. R.; Fleischhauer C.; Murray, K. Sustainability of Digital Formats: Planning for the Library of Congress Collections. URL: <https://www.loc.gov/preservation/digital/formats/> (24. svibnja 2018.)
6. Bello, J. P.; Chew, E.; Trunbull, D. ISMIR 2008: Proceedings of the 9th International Conference of Music. Philadelphia : Drexel University, dostupno i na: <https://books.google.hr/books?id=OHp3sRnZD-oC>. (31. svibnja 2018.)
7. Document management – Electronic document file format for long-term preservation : Part1: Use of PDF 1.4 (PDF/A-1) : 2005. Geneva : International Organization for Standardization.
8. GeoJP2. Wikipedia. URL: <http://fileformats.archiveteam.org/wiki/GeoJP2>. (27. lipnja 2018.)
9. JPEG File Interchange Format. Wikipedia. URL: https://en.wikipedia.org/wiki/JPEG_File_Interchange_Format. (21. lipnja 2018.)
10. Kavčič-Čolić, A. Trajno ohranjanje digitalnih virov: Koncepti in metode. // Knjižnica, 54(2010)1-2, str 99-119.
11. Library of Congress Recommended Formats Statement : 2017-2018. Washington DC : Library of Congress. Dostupno na: <https://www.loc.gov/preservation/resources/rfs/RFS%202017-2018.pdf> (24. svibnja 2018.)
12. Mackie, K. Adobe and Browser Makers Announce the End of Flash. Redmond. 25. srpnja 2017. URL: <https://redmondmag.com/articles/2017/07/25/adobe-ending-flash-support.aspx> (4. srpnja 2018.)
13. MusicXML. Wikipedia. URL: <https://en.wikipedia.org/wiki/MusicXML>. (31. svibnja 2018.)

14. Priporočeni formati : Arhiv družboslovnih podatkov. 2017. URL:
<https://www.adp.fdv.uni-lj.si/deli/postopek/priprava/formati/>. (24. svibnja 2018.)
15. Recommended formats : UK Data Service. URL:
<https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats>. (24. svibnja 2018.)
16. Runnardotter, M. On organizing for digital preservation : Doctoral thesis. Luelå : Luelå University of Technology, 2009.
17. Stančić, H., Digitalizacija. Zagreb : Zavod za informacijske studije, 2009.
18. Štebe, J.; Vipavc Brvar, I. Analiza stanja in perspektiva digitalne hrambe v arhivu družboslovnih podatkov (ADP). // Knjižnica, 55(2011)1, str. 57-86
19. Temporal compression. Webopedia. URL:
https://www.webopedia.com/TERM/T/temporal_compression.html. (27. lipnja 2018.)
20. The Archives Act, Australia, 1983., http://www7.austlii.edu.au/cgi-bin/viewdoc/au/legis/cth/consol_act/aa198398/s3.html. (12. ožujka 2018.)
21. Van Horik, R. Permanent pixels : Building blocks for the longevity of digital surrogates of historical photographs. Doctoral thesis. Delft : Technische Universiteit Delft, 2005.
22. Vogt-O'Connor, D. Reformating for Preservation and Access: Prioritizing Materials for Duplication // Conserve O Gram. 19/10(1995),
<https://www.nps.gov/museum/publications/conservoogram/19-10.pdf>. (3. travnja 2018.)
23. Web Publications. W3C Editor's Draft 23 August 2018,
<https://w3c.github.io/wpub/>. (3. rujna 2018.)
24. XML. Wikipedia. URL: <https://en.wikipedia.org/wiki/XML>. (29. lipnja 2018.)