



PAM Pokrajinski
arhiv
Maribor

Moderna arhivistika

Časopis arhivske teorije in prakse
Journal of Archival Theory and Practice

Letnik 1 (2018), št. 1 / Year 1 (2018), No. 1

Maribor, 2018

Moderna arhivistika

Časopis arhivske teorije in prakse

Journal of Archival Theory and Practice

Letnik 1 (2018), št. 1 / Year 1 (2018), No. 1

ISSN 2591-0884 (online)

ISSN 2591-0876 (CD_ROM)

Izdaja / Published by:

Pokrajinski arhiv Maribor / Regional Archives Maribor

Glavni in odgovorni urednik / Chief and Responsible editor:

*Ivan Fras, prof., Pokrajinski arhiv Maribor, Glavni trg 7, SI-2000 Maribor,
telefon/ Phone: +386 2228 5017; e-pošta/e-mail: ivan.fras@pokarh-mb.si*

Glavna urednica / Editor in chief:

mag. Nina Gostenčnik

Uredniški odbor / editorial board:

- dr. Thomas Aigner, Diözesanarchiv St. Pölten, Avstrija
- dr. Borut Batagelj, Zgodovinski arhiv Celje, Slovenija
- dr. Bojan Cvelfar, Arhiv Republike Slovenije, Slovenija
- mag. Nada Čibej, Pokrajinski arhiv Koper, Slovenija
- Ivan Fras, Pokrajinski arhiv Maribor, Slovenija
- mag. Nina Gostenčnik, Pokrajinski arhiv Maribor, Slovenija
- dr. Joachim Kemper, Institut für Stadtgeschichte Frankfurt am Main, Nemčija
- Leopold Mikec Avberšek, Pokrajinski arhiv Maribor, Slovenija
- dr. Miroslav Novak, Pokrajinski arhiv Maribor, Slovenija
- dr. Rik Opsommer, Stadsarchief Ieper - Universiteit Gent, Belgija
- Darko Rubčić, Državni arhiv u Zagrebu, Hrvaška
- dr. Izet Šabotić, Filozofski fakultet Univerziteta u Tuzli, Bosna in Hercegovina
- mag. Boštjan Zajšek, Pokrajinski arhiv Maribor, Slovenija

Recenziranje / Peer review process:

Prispevki so recenzirani. Za objavo je potrebna pozitivna recenzija. Proces recenziranja je anonimen. / All articles for publication in the conference proceedings are peer-reviewed. A positive review is needed for publication. The review process is anonymous.

Lektoriranje / Proof-reading:

mag. Boštjan Zajšek, mag. Nina Gostenčnik

Prevajanje:

mag. Boštjan Zajšek (slovenščina), mag. Nina Gostenčnik (slovenščina, angleščina), Lučka Mlinarič (bosanščina, hrvaščina, srbščina)

Oblikovanje in prelom / Design and typesetting:

mag. Nina Gostenčnik

Objavljeni prispevki so prosto dostopni. Vse avtorske pravice ima izdajatelj Pokrajinski arhiv Maribor.

©Pokrajinski arhiv Maribor Za prijavo in objavo prispevkov ni potrebno plačilo. / The publication offers open access to whole texts of the published articles. ©Pokrajinski arhiv Maribor. All articles are published free of charge.

<http://www.pokarh-mb.si/si/p/3/49/moderna-arhivistika.html>

Prejeto / Received: 20. 02. 2018

1.01 Izvirni znanstveni članek

1.01 Original scientific article

ANALIZA USTREZNOSTI FORMATOV ZA DOLGOROČNO HRAMBO IN METAPODATKOV PRI DIGITALIZACIJI ARHIVSKE KULTURNE DEDIŠČINE

dr. Gregor JENUŠ

Arhiv Republike Slovenije, Ljubljana, Slovenija
gregor.jenus@gov.si

dr. Tatjana HAJTNIK

Arhiv Republike Slovenije, Ljubljana, Slovenija
tatjana.hajtnik@gov.si

Anja PAULIČ

Arhiv Republike Slovenije, Ljubljana, Slovenija
anja.paulic@gov.si

Vanja PFAJFAR

Arhiv Republike Slovenije, Ljubljana, Slovenija
vanja.pfajfar@gov.si

Izvleček:

Digitalizacija arhivske kulturne dediščine sodi med ene od temeljnih nalog in procesov, ki so na podlagi arhivske zakonodaje zaupani slovenskim arhivom. Njena vloga se je skozi zadnja desetletja spreminjala. Od procesa, s katerim zagotavljamo dolgoročno materialno varstvo arhivskega gradiva, saj s pretvorbo gradiva iz analogne v digitalno obliko zavarujemo originalno arhivsko gradivo pred materialnimi poškodbami, se je digitalizacija razvila v proces, ki omogoča širšo diseminacijo arhivskega gradiva, saj z njo gradivo postane dostopnejše tako v obstoječih čitalnicah slovenskih arhivov kot tudi v virtualni arhivski čitalnici, ki jo slovenska javna arhivska mreža razvija v okviru projekta e-ARH.si. Avtorji se v prispevku posvečajo spremembam vloge digitalizacije, njene vrednosti v sodobni informacijsko-komunikacijski družbi kakor tudi konkretnim strokovnim vprašanjem, ki posegajo na področje priprave arhivskega gradiva za digitalizacijo, določitve metapodatkov ter formatov za dolgoročno hrambo, obsegajo pa tudi njihovo skladnost z mednarodnimi standardi.

Ključne besede:

digitalizacija, kontrolni seznam, metapodatki, formati za dolgoročno elektronsko hrambo, kontrola kvalitete, analiza

Abstract:

Suitability Analysis of Formats for Long-Term Preservation and Metadata in the Digitization of Archival Cultural Heritage

Digitization of archival records is among the core tasks and processes entrusted to the Slovenian archives by the Slovenian archival legislation. Thus, its role has changed through the decades. Digitization of archival records received a new role. From a process for a long-term preservation of archival cultural heritage, achieved by the conversion of archival records from their analog (physical) form to a digital format that protects original archival records from material damage, to a process that allows a widespread dissemination of it. Archival records become more accessible in existing reading rooms of the Slovenian archival network, as well as in the virtual archival reading room, which is currently under development within the e-ARH.si project. The article deals with changes in the role of digitization, its value in the information-communication society, as well as questions regarding the preparation of archival records for digitization, digitization formats for long-term preservation and associated metadata and their compliance with international standards.

Key words:

digitization, checklist, metadata, formats for long-term electronic preservation, quality control, analysis

1 Uvod

Digitalizacija arhivske kulturne dediščine sodi med ene od temeljnih nalog in procesov, ki so na podlagi arhivske zakonodaje¹ zaupani slovenskim arhivom. Njena vloga se je skozi zadnja desetletja spreminjala. Če je digitalizacija sprva predstavljala predvsem proces dolgoročnega zagotavljanja materialnega varstva arhivskega gradiva, ki so ga s pretvorbo iz analogne (tj. fizične oblike) v digitalno obliko zavarovali pred materialnimi poškodbami, ki so plod prepogoste ali celo neustrezne rabe ali 'stranski produkt' izdelave reprodukcij posameznih dokumentov za potrebe uporabnikov, se je njena vloga danes korenito spremenila. Zapleteni postopki virtualizacije arhivov in njihovih storitev, ki brišejo mejo med analognim in virtualnim svetom, so namreč povzročili, da je golo materialno varstvo arhivskega gradiva, čeprav ne nepomembno vodilo pri digitalizaciji, stopilo v ozadje. Digitalizacija je namreč z vidika virtualizacije arhivskega dela postala eno od ključnih orodij napredka informacijsko-komunikacijske tehnologije, ki je spremenil odnos sodobne družbe do dostopnosti informacij. Te morajo biti dostopne takoj, na daljavo (»access on demand«) in brez ovir na podlagi različnih nosilcev ali uradnih ur njenih upravljalcev. UVDAG tako digitalizacijo opredeljuje kot »/.../ pretvorbo in zajem dokumentarnega in arhivskega gradiva, izvorno nastalega v fizični ali elektronski analogni obliki (UVDAG, 13. člen).

Že leta 2015 se je *Evropska komisija v Strategiji za enotni digitalni trg* izrekla za ukrepe, ki bi omogočili učinkovitejšo uporabo digitalnih tehnologij, s katerimi bi lahko različni deležniki izrabili priložnosti, ki jih te tehnologije omogočajo.

»V kontekstu digitalnega prostora, pa naj gre za zasebni ali poslovni vidik, je ključna strateška naloga odprava omejitev, ki preprečujejo oblikovanje enotnega evropskega digitalnega trga. /.../ Tako bo oblikovan enotni evropski digitalni prostor (trg), ki bo podjetjem omogočal enostavnejše poslovanje, nižje stroške nastopanja na enotnem trgu, izboljšal bo možnosti za raziskave, razvoj in inovacije, na uporabniški strani pa bo zagotovil enake možnosti dostopa prebivalcev do digitalnih storitev, vsebin in proizvodov. Z enakimi pravili poslovanja, varnosti in zaščite se bo predvidoma okrepilo tudi zaupanje uporabnikov v digitalne tehnologije in storitve.« (Strategija, 2015).

Digitalizacija arhivskega gradiva zagotovo sodi med navedene ukrepe. Slednje izhaja tudi iz strategije *Digitalna Slovenija 2020 – Strategija razvoja informacijske družbe do leta 2020 (2016)*, v katero je Republika Slovenija pretila načela in usmeritve evropskih priporočil na področju digitalnih tehnologij in storitev.

»Digitalizacija, digitalne vsebine in e-storitve ter javna dostopnost so tudi na področju kulture bistvenega pomena za ustvarjanje dodane vrednosti znotraj same kulture ter z njo povezanih kulturnih industrij. Pomenijo osnovo za uporabo digitalnih kulturnih vsebin v procesih izobraževanja, usposabljanja, raziskovanja, vseživljenjskega učenja in pridobivanja znanja. Prispevajo k razvoju kreativnih vsebin za oblikovanje nove oziroma obogatene ponudbe ter promocije kulturne dediščine in države. Digitalne kulturne vsebine so eden najučinkovitejših instrumentov za dvig prepoznavnosti in konkurenčnosti Slovenije v Evropi in svetu.« (Digitalna Slovenija 2020, 2016).

To vodilo so slovenski arhivi povzeli v dveh strategijah: Strategija in izvedbeni načrt razvoja slovenskega elektronskega arhiva za obdobje 2010–2015 (2010) in Strategija in izvedbeni načrt razvoja slovenskega elektronskega arhiva eARH.si za obdobje 2016–

¹ Veljavna slovenska arhivska zakonodaja zajema: *Zakon o varstvu dokumentarnega in arhivskega gradiva ter arhivih (ZVDAGA; Ur. l. RS, št. 30/2006, 24/2014 - Odl. US in 51/2014); Uredba o varstvu dokumentarnega in arhivskega gradiva (UVDAG; Ur. l. RS, št. 42/2017).*

2020 (2016).² Slovenska javna arhivska služba³ (v nadaljevanju SJAS) se je še posebej v drugem dokumentu obvezala poiskati rešitve za zagotovitev učinkovite in varne digitalizacije ter njeno implementacijo v smislu povečanja učinkovitosti delovanja slovenske javne arhivske mreže. S področjem digitalizacije se tako ukvarja kar nekaj strateških ciljev projekta e-ARH.si. V okviru strateških ciljev dolgoročnega ohranjanja arhivskega gradiva in tudi povečanja dostopnosti arhivskega gradiva se projektni cilji neposredno dotikajo strategije Digitalne Slovenije 2020 in tudi usmeritev Evropske komisije. Namen projektne skupine je tako soočiti se z dediščino digitalizacije, pomeni v preteklih letih digitaliziranega arhivskega gradiva, kakor tudi določiti okvire digitalizacije za prihodnost, kar obsega celoten življenjski cikel digitalizacije od evidentiranja arhivskega gradiva do pretvorbe in zajema gradiva v digitalno obliko vključno s pripadajočimi tehničnimi, strukturnimi in vsebinskimi informacijami digitalizata, kontrolo digitalizacije, ustrezno dolgoročno hrambo digitalizata (UVDAG, 13. člen).

2 Zakonska podlaga digitalizacije in njena standardizacija

Kot že uvodoma poudarjeno, narekuje okvirje digitalizacije oz. pretvorbe arhivskega gradiva iz analogne (tj. fizične oblike) v digitalno obliko slovenska arhivska zakonodaja. ZVDAGA določa, da mora biti arhivsko gradivo pretvorjeno iz fizične v digitalno obliko na tak način, da je zagotovljena avtentičnost, celovitost in uporabnost njegove vsebine (ZVDAGA, 9.–12. člen). V teh določilih se postopki neposredno dotikajo tako imenovanih temeljnih načel ZVDAGA, ki so pri digitalizaciji še posebno pomembna, tj. načela ohranjanja gradiva in uporabnosti njegove vsebine, celovitosti, trajnosti in dostopnosti (glej poglavje: »Temeljna načela zakona«, ZVDAGA, 3.–7. člen).

Načelo **ohranjanja gradiva in uporabnosti njegove vsebine** je v primerih digitalizacije tesno povezano z avtentičnostjo pretvorjenega arhivskega gradiva. Zavedati se je treba, da digitalizacija, ne glede na to, kako strokovno je izvedena, nikoli ne more nadomestiti izvornika arhivskih dokumentov, ampak lahko dolgoročno zagotovi le varstvo informacij, ki obstajajo v njem. Z digitalizacijo podaljšamo obstoj fizičnega gradiva, slednje je namreč od takrat spravljeno v arhivskih skladiščih v boljših klimatskih pogojih (Pfajfar, 2018, str. 5–6), varno pred neustrezno rabo v čitalnicah arhivov, saj se v le-teh praviloma uporablja njegova digitalna kopija. Morebiti se komu poraja vprašanje, zakaj beseda »praviloma«. »Praviloma« zato, ker se seveda pojavljajo izjeme. V čitalnicah arhivov se pogosto najdejo posamezniki, ki izpostavljajo, da morajo arhivsko gradivo »začutiti«. Pri tem navajajo najrazličnejše razlage, zakaj je to nujno potrebno – od »skritih informacij«, ki stopijo na plano le ob podrobnem pregledu arhivskih dokumentov, do »intimnih« stikov z arhivskim gradivom, ki jih lahko doživijo le z osebnim stikom. Šalo na stran; odločitev, da je kljub digitalizaciji arhivskega gradiva in dostopnosti digitalizata uporabnikom omogočen dostop do originalov, leži v pristojnosti vodstev arhivov, a če želimo digitalizaciji dati pravo vrednost, se je tem izjemam treba izogniti.

² Če pustimo ob strani strategije, ki se nanašajo na informacijsko-komunikacijsko tehnologijo, je zanimivo, kako se je spreminjal odnos do kulturne dediščine in njene digitalizacije na ravni slovenske kulturne politike. Nacionalni program za kulturo 2004–2007 (2004) (Resolucija, Ur. l. RS 28/04) še ni predvideval digitalizacije vse kulturne dediščine, omenjal je le knjižnično in avdiovizualno gradivo, šele v naslednjem obdobju, kot zasledimo v Nacionalnem programu za kulturo 2008–2014 (2008) (Resolucija, Ur. l. RS 35/08 in 95/10), se je ta pomanjkljivost odpravila in v programu je bila izpostavljena prioriteta naloga digitalizacije kulturne dediščine, ki je pomenila »osnovo za neposredno uporabo digitalnih kulturnih vsebin v procesih izobraževanja, usposabljanja, vseživljenjskega učenja in raziskovanja«. V načrt digitalizacije so bili vključeni tudi arhivi in gradivo, ki ga slednji hranijo (Pfajfar, 2018, str. 5).

³ SJAS sestavljajo Arhiv Republike Slovenije (Arhiv RS) ter šest regionalnih arhivov (Zgodovinski arhiv Ljubljana (ZAL), Pokrajinski arhiv Maribor (PAM), Zgodovinski arhiv Celje (ZAC), Zgodovinski arhiv na Ptuj (ZAP), Pokrajinski arhiv Koper (PAK) in Pokrajinski arhiv Nova Gorica (PANG)).

Obstajajo tudi primeri, ko uporabniki digitalizaciji preprosto ne zaupajo. Leta 2011 je bilo v časniku *Dnevnik* zapisano, da »Kdor ima arhiv, ima moč«⁴. Dogodki zadnjih let so večkrat dokazali, da informacije iz arhivskega gradiva lahko predstavljajo sredstvo v medsebojnih političnih ali drugačnih obračunih. Zato dvom uporabnikov v avtentičnost digitalnih kopij ni zanemarljiv.⁵ Prav zato je treba pri digitalizaciji avtentičnosti arhivskega gradiva posvečati tolikšno pozornost. Slednja se zagotovi tako, da se ob pretvorbi izvajajo kontrole, digitaliziranim posnetkom pa dodajo tudi tehnične, strukturne in vsebinske informacije (oz. metapodatki). Metapodatki so tehnične, strukturne in vsebinske informacije, ki opisujejo tako fizično gradivo kot njegovo digitalno različico po postopku digitalizacije. Omogočajo iskanje oz. identifikacijo arhivskega gradiva, njegovo vsebinsko obdelavo, prikazovanje povezav med dokumenti, dokazovanje izvora arhivskih dokumentov in zagotavljajo celovitost (izvirnost in provenienco digitaliziranega arhivskega gradiva (ZVDAGA, 10. in 12. člen) ter avtentičnost gradiva (*Enotne tehnološke zahteve* – ETZ, II. del, 2013). Metapodatki za opis digitalne kopije omogočajo kontrolo nad kvaliteto izvedene digitalizacije in olajšajo nadaljnje upravljanje z digitalnim arhivskim gradivom (Pfajfar, 2018, str. 6).

Vsebinski ali tudi informacijski metapodatki, ki vsebujejo podatke o naslovu, vsebini, datumu nastanka, avtorju oz. pošiljatelju (ali naslovnik) ter enolično identifikacijsko oznako, so za identifikacijo digitalizata zelo pomembni in so v skladu z ETZ in standardom International Standard for Archival Description (ISAD(G), 2000), a nikakor ne zadostujejo za celovito identifikacijo digitalizata. Podatki so namreč pomanjkljivi, saj zajemajo le vsebinske podatke digitalizata. Da zagotovimo avtentičnost digitalizata ali digitalne kopije arhivskega gradiva, mora ta vsebovati tudi strukturne in tehnične podatke, ki vključujejo tudi informacije o velikosti datotek, formatih, parametrih (ločljivost, barvni model, bitna globina) in o izvedbi digitalizacije. Strukturni podatki nam dajejo informacije o relacijah, ki jih ima popisna enota (dokument) z drugimi enotami, oziroma o tem, kako je skupek dokumentov sestavljen, da tvori zaključeno enoto (Corrado in Moulaison Sandy, 2017; Pfajfar, 2018, str. 6); slednje pomeni, da so bile digitalizirane zaključene celote arhivskega gradiva, k čemur naj bi v postopkih digitalizacije arhivi tudi stremeli.

Celovitost je v postopku digitalizacije pomembna z dveh vidikov. Po eni strani je s celovitostjo mišljen konkreten postopek digitalizacije zaključenih celot, s čimer se izognemo nekontroliranemu, »ad hoc« digitaliziranju posameznih dokumentov, saj se s tem lahko izgubi njihov pomen in kontekst, namesto tega pa se osredotočimo na digitalizacijo širših in predvsem zaokroženih enot (Pfajfar, 2018, str. 6). Drugi vidik celovitosti nastopi po digitalizaciji, ko je potrebno zagotoviti nespremenljivost in neokrnjenost dobljenega dokumenta oz. reprodukcije njegove vsebine, dokazljivost izvora (provenienco) (ZVDAGA, 5., 26. in 27. člen; Hajtnik, 2016, str. 56). Digitalizirano gradivo ohrani svojo uporabnost le takrat, če je njegova vsebina enaka izvirniku in če je pretvorjena na tak način, da je omogočena **trajnost** njegovih reprodukcij ter s tem zagotovljena avtentičnost pretvorjene vsebine.

Poseben vidik digitalizacije sta **trajnost** in **dostopnost digitalizatorov**. ETZ (I. del, 2013, str. 6) opredeljujejo »poslovne, organizacijske in tehnološke pogoje« za zanesljiv zajem, pretvorbo in upravljanje z dokumenti v digitalni obliki (ne glede na to, ali so ti izvirno digitalni ali digitalizirani). Za arhivsko gradivo v digitalni obliki je treba zagotoviti dolgoročno hrambo. Ta je lahko ogrožena zaradi propadanja nosilca zapisa ali

⁴ *Kdor ima arhiv, ima moč*. V: *Dnevnik*, 19. 11. 2011. Pridobljeno 28. 3. 2018 s spletne strani: <https://www.dnevnik.si/1042425331>.

⁵ *Arhiv RS objavil tri izvornike: Dokument SDS je sestavljen iz delov različnih dokumentom*. V: *Dnevnik*, 22. 2. 2011. Pridobljeno 28. 3. 2018 s spletne strani: <https://www.dnevnik.si/1042426174>.

zastaranja tehnične opreme (ETZ, I. del, 2013). Gradivo je zato treba pretvoriti v tako obliko, ki omogoča dolgoročno hrambo (ZVDAGA, 11. člen). Pri tem se v praksi postavlja vprašanje, ali je gradivo v digitalni obliki res bolj obstojno. Problemi namreč nastanejo pri dolgoročnem ohranjanju digitalnih formatov, predvsem glede njihove uporabnosti.

A zgoraj navedeno so teoretični okviri oz. zakonski pristopi, ki jih je za optimizacijo digitalizacije potrebno prenesti v prakso.

3 Kontrolni seznam za digitalizacijo

Sledenje konkretnim delovnim procesom je pri digitalizaciji izjemno pomembno, saj samo standardizacija le-teh dejansko zagotavlja njeno pravilnost in s tem ustreznost digitalizatorov. Zakonodajalec je postopke leta 2017 podrobneje opredelil v UVDAG. Postopek digitalizacije tako mora vključevati najmanj:

»1. evidentiranje vseh enot dokumentarnega in arhivskega gradiva ne glede na obliko ali nosilec zapisa, način nastanka in druge tehnološke značilnosti;

2. pravilno pretvorbo in zajem vsebine dokumentarnega in arhivskega gradiva v digitalno obliko, ki:

- obsega vse ključne vsebinske podatke,
- zajema ali ustvari vse potrebne metapodatke, vključno s podatki, ki zagotavljajo celovitost (neokrnjenost vsebine), avtentičnost (dokazljiva povezanost reprodukcij z vsebino izvirnega dokumentarnega in arhivskega gradiva oziroma njegovim izvorom) in uporabnost dokumentarnega in arhivskega gradiva (omogočanje popolne interpretacije podatkov kot smiselne informacije z možnostjo identifikacije enot gradiva) ter omogoči strogo nadzorovano in dokumentirano dodajanje teh podatkov,
- obsega dodane podatke, ki so posebej in jasno ločeni, shranjeni in drugače označeni od izvirnih podatkov (dodatni vsebinski in tehnični metapodatki), ter vse pomembne opombe in podatke o postopku zajema in o izvornikih;

3. samodejno oziroma ročno kontrolo pravilne izvedbe digitalizacije z namenom odprave napak oziroma odstopanj;

4. evidentiranje digitaliziranega dokumentarnega in arhivskega gradiva;

5. hrambo zadostnega obsega dokumentacije, s katero se dokazuje, da uporabljena orodja, metode in postopki digitalizacije zagotavljajo zanesljiv zajem v digitalno obliko« (UVDAG, 13. člen).

Da bi se ta teoretična načela varstva arhivskega gradiva v primeru digitalizacije prenesla v prakso, so arhivi v okviru SJAS v preteklosti izdelali smernice za izvajanje digitalizacije, ki so namenjene tako ponudnikom storitev zajema arhivskega gradiva v digitalno obliko kakor tudi arhivistom, ki digitalizacijo vodijo oz. jo nadzorujejo.

Pri izbiri vrste digitalizacije se arhivisti namreč pogosto srečujejo z dilemo določitve primernih minimalnih tehničnih zahtev (parametrov) za pretvorbo v digitalno obliko. Ti sicer veljajo za enega od pomembnih dejavnikov, ki zagotavljajo kakovost izvedene digitalizacije, saj se ustvari določena stopnja zaupanja v preizkušene postopke za varno in na dolgi rok obstojno preslikavo ene oblike v drugo. Izpostaviti pa je treba, da je v teh postopkih nujno sodelovanje arhivistov, kot vsebinskih skrbnikov arhivskega gradiva, s strokovnjaki s področja informacijskih tehnologij v okviru SJAS.

V Arhivu RS se arhivsko gradivo bolj ali manj sistematično digitalizira od leta 2003, ko se je digitaliziral del gradiva katastra. Razloga za izvedbo prvega večjega projekta digitalizacije sta bila pogosto in obenem zahtevno rokovanje z gradivom (velikost katastrskih kart) ter želja uporabnikov, da pridobijo posnetke gradiva (Dobernik, 2002).⁶

Pri pripravi gradiva za digitalizacijo so arhivisti sledili določenim postopkom, ki so izhajali iz dotedanje splošne prakse in niso bili usklajeni, zapisani in torej standardizirani. Prav zato se je Arhiv RS, sledeč pobudam iz evropske *Digitalne agende*, ki je bila sprejeta leta 2010, odločil izvesti poskusno digitalizacijo z namenom priprave nacionalnih standardov za digitalizacijo arhivskega gradiva. V okviru projektov ENArC (*European Network on Archival Cooperation*) leta 2013 in v okviru projekta *Razvoj storitve sprejema in nadaljnje uporabe izbranih oblik arhivskega e-gradiva*, OPRR (*Operational Programme for Strengthening of Regional Development Potentials*) leta 2015 so bile v poskusni digitalizaciji zajete naslednje vrste arhivskega gradiva: spisovno gradivo (rokopisi, tipkopisi, tiski), fotografije, mikrofilmi, načrti, grafično in filmsko gradivo.

Kot rezultat poskusne digitalizacije, in ob sodelovanju celotne SJAS, je nastal *Kontrolni seznam za digitalizacijo arhivskega gradiva* (2016), ki velja za nekakšen akcijski dokument digitalizacijskih projektov v slovenskih arhivih. V njem so opredeljene tehnične, strukturne in vsebinske informacije, ki opisujejo tako fizično gradivo kot njegovo digitalno različico po postopku digitalizacije, vsebujejo informacije o velikosti datotek, parametrih (ločljivost, barvni model, bitna globina, format) in o izvedbi digitalizacije (*Kontrolni seznam za digitalizacijo arhivskega gradiva*, 2016).

4 Format kot eden ključnih parametrov digitalizacije

Ključni parametri digitalizacije so ločljivost skeniranja, barvna globina in format⁷. Slednji je v svetu digitalizacije, to je pretvorbe gradiva iz fizične v digitalno obliko, danes različen glede na vrsto gradiva (npr. tekstovno tiskano gradivo, kartografsko gradivo, slikovno gradivo, avdio in video gradivo, rokopisno gradivo) ali namen njegove uporabe (arhivska ali uporabniška kopija).

Datotečni format je standardni način, na katerega je informacija zakodirana v datoteki in ki določa, na kakšen način so uporabljeni biti za kodiranje informacije na digitalnem nosilcu zapisa.⁸ Čeprav je večina podatkov na svetu zapisanih v nekaj deset različnih formatih, pa so jih samo v ZDA našteji več kot 16.000 (Thomas in Martin, 2006; Hajtnik, 2016, str. 21). Bistveno je, da formati niso namenjeni katerim koli oblikam zapisov, temveč točno določenim tipom (Hajtnik, 2016, str. 21). Zato je treba dokumente, za katere je predpisano trajanje hrambe več kot pet let, iz običajne digitalne oblike pretvoriti v obliko (format), ki je primerna za dolgoročno hrambo (ZVDAGA, 11, člen). Ker pa ima vsak format dobre in slabe lastnosti, jih je treba pred izbiro za konkreten namen

⁶ Iz podobnih vzgibov so bile digitalizirane še nekatere druge zbirke, kot na primer *Zbirka listin (signatura SI AS 1063) in zemljiška knjiga plemiške posesti (t. i. Deželna deska za Kranjsko, signatura SI AS 315)* (Pfajfar, 2018, str. 12, 13).

⁷ Slovenski predpisi format poimenujejo »oblika zapisa« in jo opredeljujejo »/.../ kot organizacijske in tehnološke značilnosti zapisa, ki določajo, kako je vsebina zapisana, hranjena in prikazana v procesu hrambe« (ZVDAGA, 2. člen). Format ima več definicij. Volker Heydegger opredeljuje format kot »/.../ niz pravil, ki predstavljajo logično organizacijo podatkov in z navedbo, kako jih razlagati.« Podobno definicijo ponuja projekt InterPARES3 (2015), ki format opredeljuje kot »/.../ organizacijo podatkov znotraj datoteke, ki je običajno zasnovana za lažje shranjevanje, iskanje, obdelavo, predstavitev oz. prenos podatkov s programsko opremo«. (Hajtnik, 2016, str. 21).

⁸ File format. Pridobljeno 28. 3. 2018 s spletne strani: https://en.wikipedia.org/wiki/File_format.

hrambe poznati ter uskladiti z lastnimi zahtevami in dolgoročnim tveganjem v zvezi s temi lastnostmi (Hajtnik et al., 2009).

Formate se praviloma presoja po sledečih kriterijih⁹:

- *Uveljavljenost* (razširjenost formata): označuje stopnjo sprejetosti formata pri ustvarjalcih in uporabnikih elektronskih zapisov. Biti mora široko priznan, uveljavljen in v skupni uporabi na širšem področju, pomembna je tudi njegova podpora s strani arhivskih institucij.
- *Interoperabilnost* (odvisnost formata): do kolikšne mere je format odvisen od strojne oziroma programske opreme oz. operacijskih sistemov ter kakšna je posledično kvaliteta dokumentiranosti.
- *Odprtost formata*: do kolikšne mere je specifikacija formata dostopna javnosti (lastniški, nelastniški format)¹⁰.
- *Kakovost* (transparentnost formata): do kolikšne mere lahko preverjamo strukturo in attribute formata, možnost pretvorbe vsebine v obliko, ki je neposredno uporabna (ponovna uporaba), vključenost mehanizmov za zaščito.
- *Samodokumentiranost* (podpora za metapodatke): do kolikšne mere so podprti metapodatki – osnovni opisni metapodatki kot tudi tehnični in upravni metapodatki, ki se nanašajo na kreiranje in zgodnje faze življenjskega cikla dokumenta (Hajtnik, 2016, str. 165–168).

⁹ V okviru raziskave *Celovit pristop k pretvorbi elektronskih dokumentov v obliko za dolgoročno hrambo* je Tatjana Hajtnik izvedla raziskavo *datotečnih formatov*. Raziskava je formate razvrstila glede na njihov razvoj, dostopnost in vzdrževanje specifikacij, razvijalce oz. proizvajalce ter jakost vpliva formata na morebitno izgubo bistvenih lastnosti e-dokumentov. V okviru raziskave, ki je zaobjela skupno 22 držav oz. 24. anketirancev (institucij) iz EU in zunaj nje, je bila izvedena tudi anketa glede datotečnih formatov za dolgoročno hrambo, ki jih institucije uporabljajo. Pridobljeni podatki so bili analizirani in glede na njihove opise v literaturi sintetizirani v model lastnosti formatov za dolgoročno hrambo e-dokumentov »MkiFormat« (Hajtnik, 2016).

¹⁰ Omejitve pri rokovanju s formati nastopajo glede razvoja, dostopnosti in vzdrževanja specifikacij formatov. Iz tega vidika formate delimo na nelastniške in lastniške.

Nelastniške formate (angl. non-proprietary) pogosto imenujemo kar odprti formati (angl. open formats). Podprti so s strani več razvijalcev in jih je mogoče uporabljati z različno programsko opremo. Te formate običajno sprejemajo in nato upravljajo konzorciji zainteresiranih organizacij (npr. Organization for the Advancement of Structured Information Standards - OASIS, World Wide Web Consortium – W3C) ali pa nacionalna (npr. Slovenski inštitut za standardizacijo - SIST) in mednarodna standardizacijska telesa (npr. International Organization for Standardization - ISO). Odprte formate lahko uporabi in implementira kdor koli ter brez kakršnih koli omejitev (licenca ni plačljiva, ni lastništva specifikacije). Specifikacije so prosto dostopne, kar omogoča njihove implementacije in prilagoditve v različnih okoljih. Organizacije in posamezniki, ki svoje dokumente shranjujejo v odprtih formatih, se lahko izognejo odvisnosti od samo enega proizvajalca oz. razvijalca programske opreme. Na ta način lahko programsko opremo prosto zamenjajo, če npr. proizvajalec zviša njeno ceno ali programsko opremo spremeni, neugodno spremeni licenčno politiko ali pa programsko opremo preneha prodajati ali vzdrževati.

Lastniški formati (angl. proprietary) so pod nadzorom in podpora samo enega proizvajalca oz. razvijalca programske opreme. Specifikacije teh formatov niso javno dostopne, njihova uporaba je pogosto vezana na licenčne pogoje (običajno plačljiva) lastnika izvorne kode formata. Pri uporabi lastniških formatov se lahko zgodi, da v nekem trenutku starih dokumentov ni več mogoče odpreti ali pa je lahko ta pretvorba neuspešna. E-dokumenti za predstavitev vsebine vsebujejo kodirne znake. Vsak znak je zapisan s kodo določenega števila bitov. Kode pomenijo črke, številke in druge znake, ki se pojavljajo v besedilih. Znaki so urejeni s pomočjo kodnih tabel, ki povezuje grafično predstavitev nekega znaka z njegovim binarnim zapisom. Med bolj znanimi kodnimi tabelami so npr. tabela ASCII (tudi ISO standard), Extended ASCII (iz katerega so nastale npr. ANSI, ISO Latin-1 in Latin-2), Unicode (standard ISO 10646). Zaradi svoje preprostosti in standardiziranosti zagotavljajo vsesplošno berljivost ter so neodvisni od strojne in programske opreme, zato so primerni tudi za dolgoročno e-hrambo (npr. primer format TXT). S pretvorbo zahtevnejših formatov v take preproste formate lahko rešujemo tudi težave združljivosti (Hajtnik, 2016, str. 22, 23).

Zavedati se moramo, da so formati podvrženi hitremu tehnološkemu razvoju, kar pomeni, da poznamo znotraj posameznih formatov veliko različic. Da je temu tako, nam potrjuje že hiter pogled v register PRONOM¹¹, iz katerega je razvidno, obstaja format .pdf v 37 unikatnih identifikacijskih številkah fmt in format .tiff v šestih, različice se razlikujejo glede na razvoj standarda, s tem povezano specifikacijo in namen uporabe (Hajtnik, 2016, str. 169).

Direktiva EU (Direktiva 2013/37/EU) iz leta 2013 določa, da morajo biti dokumenti na voljo v odprtem in strojno berljivem formatu skupaj z njihovimi metapodatki.¹² Med formati, s katerimi se srečujejo arhivi v svojih postopkih digitalizacije najpogosteje, srečamo formate JPEG/JPG (*Joint Photographic Experts Group*), TIFF (*Tagged Image File Format*), JPEG2000 in PDF (*Portable Document Format*)¹³, z izjemo avdio-video gradiva, kjer prevladujejo formati MPEG (*Moving Picture Experts Group*), WAVE (*Waveform Audio File Format*) in FLAC (*Free Lossless Audio Codec*).

S stališča arhiviranja, dolgoročne hrambe oziroma ustvarjanja uporabniških kopij so nekateri formati boljši od drugih, njihova poglobljena dobra lastnost pa je točen prikaz vsebine oziroma shranjenih informacij skozi čas, ne glede na razvoj v strojni in programski opremi. Primernost formata, ki ga izberemo za dolgoročno hrambo ali za uporabniško kopijo, lahko ocenjujemo po različnih kriterijih. V glavnem si vse ustanove, kot so arhivi in knjižnice, ki se ukvarjajo z digitalizacijo kulturne dediščine in ponujanjem digitalnih vsebin uporabnikom, določijo nabor lastnosti, ki naj bi jih digitalizirano ali izvorno digitalno gradivo imelo. Na tej podlagi nato naročijo ali izvedejo digitalizacijo ali pretvorbo gradiva v ustrezen format oziroma formate. Izbira formata je v prvi vrsti odvisna od vrste gradiva (npr. spisovno, avdiovizualno, fotografsko) in od predvidenega načina uporabe gradiva. Format, primeren za dolgoročno hrambo, je namreč lahko popolnoma neprimeren za prikaz in objavo gradiva na spletu ali za izdelavo uporabniške kopije.

Izbira ustreznega datotečnega formata je eden od najpomembnejših korakov pri postopku digitalizacije arhivskega gradiva, česar se v slovenski javni arhivski mreži vse bolj zavedamo. Poznavanje širokega nabora formatov za raznovrstno arhivsko gradivo zahteva sodelovanje informatikov, arhivistov in specialistov za posamezno vrsto arhivskega gradiva. Upravljanje s formati mora postati del strategije za materialno varovanje arhivskega gradiva in hkrati osnova za pokrivanje potreb različnih tipov uporabnikov (Digital Preservation Handbook, 2015). Kot del projekta *Razvoj storitve sprejema in nadaljnje uporabe izbranih oblik arhivskega e-gradiva* smo v slovenski javni arhivski mreži določili izhodne formate digitalizacije, ki jih izvajalcu digitalizacije, skupaj z drugimi zahtevami, posredujemo v tako imenovanem kontrolnem seznamu, dokumentu, ki določa parametre digitalizacije. Za upravljanje z gradivom v arhivu

¹¹ Register PRONOM je prvotno za svoje potrebe razvil Nacionalni arhiv Velike Britanije (*The National Archives*), danes pa je na voljo vsem, ki potrebujejo zanesljiv in verodostojen informacijski vir o formatih ter programski opremi, njenih življenjskih ciklih in tehničnih specifikacijah, ki jih podpira. Vsebuje podatke o praktično vseh poznanih formatih na svetu ter pripadajoče opise (Hajtnik, 2016, str. 62).

Glej tudi spletni vir (pridobljeno 28. 3. 2018): <http://www.nationalarchives.gov.uk/PRONOM/BasicSearch/proBasicSearch.aspx?status=new>.

¹² Pod strojno berljiv format direktiva opredeljuje datotečni format, ki je strukturiran tako, da lahko programske aplikacije z lahkoto določajo, prepoznavajo in izločajo določene podatke, vključno s posameznimi dejanskimi navedbami in njihovo notranjo strukturo.

Pod izrazom odprti format direktiva opredeljuje datotečni format, ki je neodvisen od uporabljene platforme in javnosti dostopen brez omejitev, ki bi preprečevala ponovno uporabo dokumentov (Hajtnik, 2016, str. 59.)

¹³ Format PDF se priporoča za dolgoročno hrambo besedilnih dokumentov, medtem ko se format TIFF (ISO 12639, ver. 6) priporoča za grafične dokumente (podrobneje glej: ETZ 2.1, I del, 2013; Hajtnik, 2016, str. 42, 43).

potrebujemo datoteko v dveh kopijah, in sicer kopijo v formatu za dolgoročno hrambo in v formatu za ogledno oziroma uporabniško kopijo (Kontrolni seznam, 2014).

5 Predstavitev nekaterih formatov, ki se uporabljajo pri digitalizaciji v SJAS

Vodilo pri izbiri formatov naj bo doseči optimalno rešitev v okviru danih finančnih in tehnoloških zmožnostih. Podpora prevelikemu številu formatov ni ekonomična, saj za seboj potegne večjo potrebo po diskovnih kapacitetah, programske opreme za prikazovanje različnih formatov in ne nazadnje vnaša zmedo v sam postopek digitalizacije (Digital Preservation Handbook, 2015). V nadaljevanju povzemamo poglobitve značilnosti datotečnih formatov, za katere se trenutno odločamo pri digitalizaciji arhivskega gradiva in so primerni bodisi za dolgoročno hrambo ali za izdelavo uporabniških oziroma oglednih kopij. Skupne značilnosti opisanih datotečnih formatov so velika razširjenost, dobra podpora in konstanten razvoj. Vsi formati imajo razširjene in lahko dostopne pregledovalnike.

TIFF

Format TIFF je široko razširjen, fleksibilen in dobro podprt datotečni format za shranjevanje in izmenjavo rastrskih slik¹⁴. Podpira več tipov kompresij ter nekomprimiranih datotek in barvne globine od 1 do 24 bitov. Specifikacija formata TIFF je v lasti družbe Adobe, vendar je za uporabnike prosto dostopna. V format je vključena celostna podpora za metapodatke, mogoče pa je dodajati lastne označevalce.¹⁵ Nekomprimiran format TIFF je prednostni in brezizgubni format za dolgoročno hrambo, saj je datoteke v tem formatu mogoče urejati in shranjevati brez izgubljanja kvalitete. Slabost formata je znatna velikost datotek (ena datoteka je lahko velika tudi do 4 GB), kar otežkoča delo s samimi datotekami, ki poleg tega porabijo precej prostora na diskovju. Omeniti velja še podtip formata TIFF GeoTIFF¹⁶, ki omogoča georeferenciranje rastrskih slik in je format, primeren za dolgoročno hrambo geopodatkov. GeoTIFF je tudi dober primer fleksibilnosti formata TIFF, ki se kaže v dodajanju lastnih oznak in s tem razširjanju metapodatkovnih elementov. Format GeoTIFF za dolgoročno hrambo prostorskih podatkov uporabljajo v *Danskem nacionalnem arhivu* (Hajtnik, 2016, str. 76–77).

JPEG

Tudi format JPEG je zelo razširjen rastrski slikovni format. V nasprotju s formatom TIFF za kodiranje uporablja izgubno kompresijo, kar pomeni, da se določene informacije v sliki lahko izgubijo. Stopnjo kompresije lahko prilagajamo in s tem lovimo ravnotežje med velikostjo datoteke in kvaliteto slike.¹⁷ Že manipuliranje z datoteko JPEG (preimenovanje, premikanje, ponovno shranjevanje) povzroča izgubljanje kvalitete, zato format ni primeren za izdelavo arhivske kopije,¹⁸ pogosto pa ga uporabimo kot format za

¹⁴ Rastrska slika je dvodimenzionalna mreža pikselov, v kateri ima vsak piksel svojo barvno vrednost.

¹⁵ Več: *Graphic File Formats* (2008). Pridobljeno 28. 3. 2018 s spletne strani: <http://www.nationalarchives.gov.uk/documents/information-management/graphic-file-formats.pdf>.

¹⁶ Več: *GeoTIFF, Revision 1.0* (2011). Pridobljeno 28. 3. 2018 s spletne strani: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000279.shtml>.

¹⁷ Več: *JPEG*. Pridobljeno 28. 3. 2018 s spletne strani: <https://en.wikipedia.org/wiki/JPEG>.

¹⁸ Več: *Best Image Formats for Archival Photography* (2016). Pridobljeno 28. 3. 2018 s spletne strani: <https://digitalarch.org/blog/2017/4/7/best-image-formats-for-archival-photography>.

uporabniško kopijo pri kartografskem in drugem slikovnem gradivu.¹⁹ Pri pretežno tekstovnem gradivu za uporabniško kopijo raje izberemo format PDF, ki omogoča lažje pregledovanje oziroma listanje.

PDF

Nesporna zmagovalka med formati za dolgoročno hrambo spisovnega gradiva v svetu je različica formata PDF – PDF/A. Tako ne preseneča, da tudi v Arhivu RS ter ostalih članicah SJAS prevladuje PDF kot eden od formatov, v katerem hranimo digitalizirano arhivsko gradivo. Format PDF je v zgodnjih devetdesetih letih ustvaril sistem *Adobe Systems*, ker je potreboval izmenjavo dokumentov v elektronski obliki med svojimi zaposlenimi, tehnologije pa tega takrat niso omogočale. Format ni bil namenjen urejanju vsebine, temveč le prikazovanju in ohranjanju »videza in obnašanja« (angl. »*look and feel*«). Njegov ključen namen je bil, da bi bili elektronski dokumenti v bistvu podobni papirnim ekvivalentom in verodostojni, zanesljivi ter enostavni za uporabo (Hajtnik, 2016, str. 68).

Skozi razvoj interneta in uporabo PDF je ta postal de facto standard za izmenjavo dokumentov vključno z vektorsko grafiko (ilustracije in modeli), besedilom in rastrsko grafiko (fotografije in druge vrste slik). Mnoge študije so pokazale, da format PDF ni primeren za dolgoročno hrambo glede na svoje kriterije, zato je *Adobe Systems Inc.* specifikacijo PDF večkrat spremenil, leta 2005 pa objavil prvo različico formata PDF/A, ki ga je z leti dopolnjeval. PDF/A je standardizirana verzija formata PDF (družina standardov ISO 19005). Od PDF se razlikuje v tem, da so odstranjene značilnosti, ki so neprimerne za dolgoročno hrambo, npr. povezovanje na pisave (fonte) je nadomeščeno z zahtevo po vključenosti uporabljenih fontov v sam dokument. Standard PDF/A identificira »profil« za elektronske dokumente, ki bo zagotavljal, da bodo lahko za ves čas njihove hrambe reproducirani vedno na natančno enak način. Ključni element reprodukcije je zahteva, da so dokumenti PDF/A stoodstotno neodvisni. Vse informacije, ki so potrebne za prikaz dokumentov vedno na enak način, morajo biti vgrajene v datoteko. To vključuje (ni pa omejeno na) informacijo o vsebini (besedilo, rastrske slike in vektorske grafike), pisavah in barvah. Dokument PDF/A ne sme biti odvisen od podatkov iz zunanjih virov (npr. fontov, programa in hiperpovezav). Zahteve standarda ISO za programsko opremo, ki prikazuje format PDF/A, so, da vključuje smernice za upravljanje z barvami, podporo za vgrajene pisave (fonte) in uporabniške vmesnike za branje vgrajenih opomb (npr. podčrtavanje ali označevanje besedila). Nadalje zahteva, da je barvna shema določena v obliki, neodvisni od strojne in programske opreme, metapodatki pa morajo biti zapisani v formatu XMP²⁰ (Hajtnik, 2016, str. 68–69).

Razvoj arhivskega formata PDF je spodbudilo dejstvo, da so javne ustanove producirale ogromne količine tekstovnih dokumentov, ki so jih mnogokrat hranile v formatu PDF. Vendar se PDF ni izkazal kot format, ki je primeren za arhiviranje digitalnega gradiva. Ravno na primeru formata PDF lahko najdemo številne različice, ki imajo vsaka svoj namen, vendar pa je vsaka tudi s svojimi prednostmi in slabostmi. Tega se žal premalo zavedamo; najpogosteje se namreč zadovoljimo s tem, da rečemo, da želimo hraniti gradivo v formatu .pdf, ker je to v svetu pač format, ki velja za primernega za dolgoročno hrambo (Hajtnik, 2016, str. 68–69). Vendar vsak format PDF ni primeren za dolgoročno hrambo arhivskega gradiva.

¹⁹ Več: *JPEG Image Encoding Family* (2012). Pridobljeno 28. 3. 2018 s spletne strani: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000017.shtml>.

²⁰ XMP (Extensible Metadata Platform) specifikacija temelji na XML standardu in omogoča uporabo shem za opis metapodatkov.

6 Format PDF ni enak formatu PDF

Kot smo že omenili, je v registru PRONOM možno za format PDF najti kar 37 unikatnih različic, ki se med seboj razlikujejo glede na razvoj standarda, s tem povezano specifikacijo in glede na namen uporabe (Hajtnik, 2016, str. 169).

PDF/A se je v preteklih letih po svetu uveljavil kot format za dolgoročno hrambo gradiva, ki zagotavlja možnost reproduciranja dokumentov v prihodnosti ne glede na programsko opremo. Dokumenti v tem formatu naj bi bili karseda samozadostni, samoopisni, njihovo prikazovanje pa neodvisno od strojne ali programske opreme (PDF/A-1, 2010). Podtip PDF/A-1a, kjer 1 pomeni prvi dodatek k standardu (obstajata še 2 in 3, v letu 2018 pa pričakujemo izid 4, ki bo slonel na nadgrajenem formatu PDF 2.0²¹), a pa nivo skladnosti s standardom (obstaja še nivo B), je format, ki trenutno pokriva zahteve dolgoročne hrambe spisovnega gradiva, hkrati pa omogoča večjo dostopnost gradiva. PDF/A-1a omogoča enako kot PDF/-1b zvesto reprodukcijo dokumenta, hkrati pa z dodatnimi posebnostmi, kot sta na primer izbira jezika in označevanje besedila²², poveča dostopnost dokumenta. Omogočanje dostopnosti pa je ključno vsaj na dveh področjih, in sicer pri omogočanju strojne berljivosti dokumentov in omogočanju dostopa do gradiva za slepe in slabovidne uporabnike. Dober dokument v PDF/A-1a je namreč skladen tudi s standardom za dostopnost PDF/UA (*Accessibility, b. d.*), prav tako je primeren za izdelavo uporabniške kopije (Hajtnik, 2016, str. 68–69).

Format PDF/A-1

PDF/A-1 je bil leta 2005 objavljen kot ISO 19005-1. Temelji na PDF 1.4 in ima omejitve glede uporabe barv, pisav, pripomb in drugih elementov. Obstajata dve ravni PDF/A-1:

- raven b (PDF/A-1b; b kot »*basic*« oz. osnovna raven) in
- raven a (PDF/A-1a; a kot »*accessible*« oz. dostopna raven).

PDF/A-1b zagotavlja, da je vizualni izgled dokumenta mogoče dolgoročno ohraniti. Skladnost s stopnjo a (PDF/A-1a) temelji na ravni b, vendar dodaja bistvene lastnosti označenega PDF: zahteva strukturne informacije in zanesljivo besedilno semantiko Unicode²³, da se ohrani logična struktura dokumenta in naravnega branja. PDF/A-1a ne zagotavlja le ohranitve vizualnega izgleda, temveč tudi, da se njegova vsebina (semantika) zanesljivo razlaga in je dostopna ranljivim skupinam (npr. slepi uporabniki lahko s programom za branje preberejo označene dokumente v PDF) (Hajtnik, 2016, str. 70).

Format PDF/A-2

Leta 2011 je bil objavljen PDF/A-2 kot ISO 19005-2, ki temelji na PDF 1.7. Vključuje številne funkcionalnosti, ki v PDF/A-1 niso na voljo, npr. stiskanje JPEG 2000, transparentnost in uporaba slojev, vgraditev datotek PDF/A. Podobno kot PDF/A-1 ponuja PDF/A-2 raven skladnosti b in a. Dodaja še nivo u, vmesno raven med ravnema PDF/A-2a in PDF/A-2b. Zahteva namreč zanesljivo semantiko Unicode, ne pa informacij o strukturi. PDF/A-2u zagotavlja, da se lahko dokumenti vizualno reproducirajo, besedilo

²¹ Več o tem: PDF/A. (b. d.).

²² Z označevanjem (tagging) dokumenta opišemo logično strukturo dokumenta (naslovi, tabele, opombe).

²³ Unicode je 16-bitni nabor znakov, ki naj bi pokrili potrebe za zapisovanje večine svetovnih naravnih jezikov. Ujema se s standardom ISO 10646 in je bil razvit pod okriljem organizacije Unicode Consortium.

se lahko ekstrahira in omogočeno je iskanje. Z vidika arhiviranja e-sporočil s priponkami je zanimiva možnost vstavljanja drugih veljavnih datotek PDF/A. Dokumenti PDF/A-2 lahko vsebujejo datoteke, če priloženi dokumenti ustrezajo ravni PDF/A-1 ali PDF/A-2 (Hajtnik, 2016, str. 70–71).

Format PDF/A-3

PDF/A-3 je bil objavljen leta 2012 kot ISO 19005-3 in podpira skladnostne ravni a, b in u. PDF/A-3 z vidika digitalizacije papirnega gradiva v digitalno ni zanimiv, saj se razlikuje od PDF/A-2 le v enem, vendar izjemno pomembnem področju – omogoča vstavljanje dokumentov poljubne vrste datotek (vključno z XML, CSV, CAD, slikami, binarno izvedljivimi datotekami), vendar pa ti ostajajo v svoji izvorni obliki, v kateri so bili ustvarjeni (Hajtnik, 2016, str. 73).

7 Identifikacija formatov

Ljudsko izročilo: »*Ni vse zlato, kar se svetil*«, lahko parafraziramo tudi na primeru formatov. Osnovno možnost identifikacije formata nam ponuja prepoznavanje končnice imena datoteke, npr. končnica .pdf naj bi pomenila, da je dokument v formatu PDF. Vendar ta način identifikacije formata še zdaleč ni zanesljiv, saj lahko končnico imena datoteke zelo enostavno in brez večjih naporov (namerno ali nenamerno) spremenimo.²⁴ Obstajajo pa tudi druge, zelo učinkovite možnosti prepoznavanja formatov, in sicer jih lahko identificiramo s pomočjo ustrezne programske opreme. Te v dokumentu poiščejo notranje karakteristike (npr. zaporedje bitov, ki se enolično pojavi le v določenih formatih) in se povezujejo za te potrebe s posebej ustvarjenimi registri formatov. Taki registri omogočajo identifikacijo in validacijo formatov ter ponujajo informacije o možnih pretvorbe, stopnji tveganja, karakterističnih lastnostih in o načinu prikazovanja.

Za potrebe prispevka in analize digitalizatorov smo uporabili prosto dostopen register formatov PRONOM²⁵ in programsko orodje DROID²⁶ (*Digital Record Object Identification*), ki ga je, prav tako kot PRONOM, razvil britanski *The National Archives* in ki avtomatsko identificira format dokumenta (ali paket dokumentov) še pred dostopom do dokumenta oz. njegovim odpiranjem. DROID je brezplačna in odprtokodna programska oprema, ki je na voljo v skladu z licenco BSD²⁷. Izvorno kodo programskega orodja DROID lahko prenesemo s portala GitHub²⁸.

²⁴ Poleg tega smo že omenili, da poznamo različne verzije formata PDF, kar pa iz samega imena datoteke ne moremo razbrati (op. a.).

²⁵ Več: PRONOM. Pridobljeno 30. 3. 2018 s spletne strani: <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>.

²⁶ Več: DROID. Pridobljeno 30. 3. 2018 s spletne strani: <http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>

²⁷ Licenca BSD je skupina licenc za programsko opremo, ki vzpostavlja minimalne omejitve glede ponovne distribucije programske opreme, ki jo pokriva. Izvorno je bila licenca uporabljena za sistem Berkeley Software Distribution (BSD), ki je podoben Unixu.

Več o licenci BSD lahko najdemo spletni strani: <http://www.lininfo.org/bsdlicense.html>. Pridobljeno 30. 3. 2018.

²⁸ GitHub je spletni gostiteljski servis za repozitorije v Git. Git je od svojega nastanka leta 2005 postal standard med sistemi za nadzor različic (VCS – Version Control System) oz. sprememb dokumentov ali kode. Omogoča sodelovanje več razvijalcev na istem projektu, obenem pa nam kadarkoli omogoča obnovitev kode na prejšnjo – zgodnejše stanje. V primeru sodelovanja več razvijalcev na istem projektu lahko kadarkoli vidimo, kdo je kaj delal, in posledično veliko hitreje rešimo probleme, ki se znajo pojaviti. V nasprotju z Git-om, ki je ukazno orodje, ponuja GitHub spletni grafični vmesnik. Zagotavlja tudi kontrolo dostopa in številne storitve za kolaborativni razvoj, npr. sledenje napakam, zahteve za lastnosti, upravljanje z nalogami in wiki za vsak projekt.

V jeseni leta 2017 je bilo uradno objavljeno programsko orodje *veraPDF* za validacijo formata PDF/A. Orodje preveri in potrdi (validira) skladnost dokumentov PDF s formatom PDF/A in omogoča validacijo vseh različic datotek PDF/A – PDF/A-1, PDF/A-2 in PDF/A-3. *veraPDF* je odprtokodni validator PDF, ki je zasnovan tako, da ustreza potrebam elektronske hrambe. Na voljo je kot grafični uporabniški vmesnik med orodjem in programsko opremo (angl. *Graphical user interface*, GUI), vmesnik z ukazno vrstico (angl. *Command-line interface*, CLI) ali kot spletna aplikacija. Na voljo je tudi javanska knjižnica za implementacijo funkcionalnosti orodja v lastne programske rešitve. Razvoj orodja je financiran s strani projekta Evropske unije PREFORMA²⁹, podpirajo ga vodilni člani skupnosti razvijalcev programske opreme PDF (več o razvoju orodja *veraPDF* glej Hajtnik, 2016, str. 65, 101).

Analiza digitalizatorov

V nadaljevanju smo izbrali šest (6) naključnih digitalizatorov iz Arhiva RS iz različnih fondov. Ustvarjeni so bili v različnih obdobjih in s strani različnih ponudnikov spremljevalnih storitev pretvorbe arhivskega gradiva iz analogne (tj. fizične) oblike v digitalno. Digitalizate smo analizirali tako, da smo najprej poiskali podatke o posameznem digitalizatu iz lastnosti datoteke, nato smo ga preverili z orodjem DROID (Slika 1), v nadaljevanju pa izvedli še validacijo formata PDF/A s programskim orodjem *veraPDF*.

Vhodni podatki so bili sledeči (tabela 1):

Tabela 1: Vhodni podatki za analizo digitalizatorov

Ime datoteke (digitalizata)	Arhivski fond	Leto digitalizacij e	Uporabniška zahteva ob naročilu digitalizacije glede formata
01	SI AS 315, Deželna deska za Kranjsko	2009	Format: PDF
126_II-3u	SI AS 1074, Zbirka urbarjev	2014 (OP RR)	Format (uporabniške kopije): / Ločljivost: / Globina: /
SI AS 1071_273	SI AS 1071, Zbirka grbov in zastav	2016	Format (arhivske kopije): PDF/A-1a Ločljivost: 600 dpi Globina: 24 bit
SI_AS_309_A_009	SI AS 309, Zbirka zapuščinskih inventarjev Deželnega sodišča v Ljubljani	2014 (OP RR)	Format (uporabniške kopije): PDF/A-1b Ločljivost: 120 dpi Globina: 24 bit
SI_AS_R_1850_004	SI AS 1850, Enaintrideseta divizija narodnoosvobodilne vojske in partizanskih odredov Jugoslavije	2014 (OP RR)	Format (uporabniške kopije): PDF/A-1b Ločljivost: 120 dpi Globina: 24 bit
SI_AS_309_F_001	SI AS 309, Zbirka zapuščinskih inventarjev Deželnega sodišča v Ljubljani	2017 (redna digitalizacija)	Format (uporabniške kopije): PDF/A-1a Ločljivost: 300 dpi Globina: 24 bit

²⁹ Več o projektu glej spletno stran: <http://www.preforma-project.eu/>. Pridobljeno 30. 3. 2018.

Slika 1: Identifikacija formata datotek z orodjem DROID

Resource	Extension	Size	Last modified	Ids	Format	Version	Mime type	PUID	Method	Hash
D:\Analiza_pdf\01.pdf	pdf	125,3 MB	28.2.2013 14:43		Acrobat PDF 1.6 - Portable Document Format	1.6	application/pdf	fmt/20	Signature	7ca4baa320346f8b3cb5bd341622288
D:\Analiza_pdf\126_II-3u.pdf	pdf	161 MB	2.2.2015 14:30		Acrobat PDF 1.4 - Portable Document Format	1.4	application/pdf	fmt/18	Signature	5a686cf0df27a5033a9b36138e4ebc6
D:\Analiza_pdf\SI AS 1071_273.pdf	pdf	17,2 MB	18.11.2016 4:07		Acrobat PDF 1.4 - Portable Document Format	1.4	application/pdf	fmt/18	Signature	a36323a07e9b03634d98e3ab643d9314
D:\Analiza_pdf\SI AS_309_A_009.pdf	pdf	7,7 MB	6.12.2014 11:05		Acrobat PDF 1.4 - Portable Document Format	1.4	application/pdf	fmt/18	Signature	cf20feb2a60ab20f69dc32b910ae912f
D:\Analiza_pdf\SI AS_309_F_001.pdf	pdf	41,6 MB	27.11.2017 8:27		Acrobat PDF 1.4 - Portable Document Format	1.4	application/pdf	fmt/18	Signature	18f7e5f1ac5571b09d6d5e85c5f72464
D:\Analiza_pdf\SI AS_R_1850_004.pdf	pdf	3,1 MB	4.12.2014 17:36		Acrobat PDF 1.4 - Portable Document Format	1.4	application/pdf	fmt/18	Signature	4cb35fb88406730567d3259900b6a553

V nadaljevanju smo za vse datoteke validirali (avtomatsko, angl: *avto-detect*) format s programskim orodjem *veraPDF* (tabela 2).

Tabela 1: Validacija formata s programom veraPDF

Ime datoteke (digitalizata)	Uporabniška zahteva ob naročilu digitalizacije glede formata	Rezultat validacije
01	PDF	Datoteka ni v nobeni različici formata PDF/A
126_II-3u	PDF/A-1b	PDF/A-1b
SI AS 1071_273	PDF/A-1a	PDF/A-1a
SI AS_309_A_009	PDF/A-1b	PDF/A-1b
SI AS_R_1850_004	PDF/A-1b	PDF/A-1b
SI AS_309_F_001	PDF/A-1a	PDF/A-1a

Tabela 2: Identifikacija formata in pregled lastnosti izbranih digitalizatorjev iz lastnosti datoteke in z orodjem DROID

Datoteka .pdf (digitalizat)	Fond	Leto digitalizacije	Lastnosti datoteke		DROID (Slika 1)	
01	SI AS 315, Deželna deska za Kranjsko	2009	Ustvarjeno:	5. 6. 2009	Ustvarjeno:	/
			Spremenjeno:	28. 2. 2013	Spremenjeno:	28. 2. 2013
			Velikost:	125,34 MB	Velikost:	125,3 MB
			PDF različica:	1.6 (Acrobat 7.x)	PDF različica:	Acrobat PDF 1.6 (fmt/20)
			Aplikacija:	Adobe Acrobat 9.1		
			Producent PDF-a:	Adobe Acrobat 9.1 Image Conversion Plug-in		
			Strani:	285		
126_II-3u	SI AS 1074, Zbirka urbarjev	2014	Ustvarjeno:	29.12.2014	Ustvarjeno:	/
			Spremenjeno:	2. 2. 2015	Spremenjeno:	2. 2. 2015
			Velikost:	161,02 MB	Velikost:	161 MB
			PDF različica:	1.4 (Acrobat 5.x) (PDF/A -1B)	PDF različica:	Acrobat PDF 1.4 (fmt/18)
			Aplikacija:	Adobe Acrobat 9.1		
			Producent PDF-a:	ABBYY FineReader 12		
			Strani:	258		
SI AS 1071_273	SI AS 1071, Zbirka grbov in zastav	2016	Ustvarjeno:	18. 11. 2016	Ustvarjeno:	/
			Spremenjeno:	18. 11. 2016	Spremenjeno:	18. 11. 2016
			Velikost:	17,17 MB	Velikost:	17,2 MB
			PDF različica:	1.4 (Acrobat 5.x) (PDF/A -1A)	PDF različica:	Acrobat PDF 1.4 (fmt/18)
			Aplikacija:	LuraDocument PDF Compressor		
			Producent PDF-a:	Sever 6.2.92.61, LuraDocument PDF v2.60		
			Strani:	2		
SI_AS_309_A_009	SI AS 309, Zbirka zapuščinskih inventarjev Deželnega sodišča v Ljubljani	2014	Ustvarjeno:	6. 12. 2014	Ustvarjeno:	/
			Spremenjeno:	6. 12. 2014	Spremenjeno:	6. 12. 2014
			Velikost:	7,73 MB	Velikost:	7,7 MB
			PDF različica:	1.4 (Acrobat 5.x) (PDF/A -1B)	PDF različica:	Acrobat PDF 1.4 (fmt/18)
			Aplikacija:	LuraDocument PDF Compressor		
			Producent PDF-a:	Sever 6.2.92.61 LuraDocument PDF v2.60		
			Strani:	7		

SI_AS_R_1850_004	SI AS 1850, Enaintrideseta divizija narodnoosvobodilne vojske in partizanskih odredov Jugoslavije	2014	Ustvarjeno:	4. 12. 2014	Ustvarjeno:	/
			Spremenjeno:	4. 12. 2014	Spremenjeno:	4. 12. 2014
			Velikost:	3,11 MB	Velikost:	3,1 MB
			PDF različica:	1.4 (Acrobat 5.x) (PDF/A -1B)	PDF različica:	Acrobat PDF 1.4 (fmt/18)
			Aplikacija:	LuraDocument PDF Compressor		
			Producent PDF-a:	Srever 6.2.92.61 LuraDocument PDF v2.60		
			Strani:	6		
SI_AS_309_F_001	SI AS 309, Zbirka zapuščinskih inventarjev Deželnega sodišča v Ljubljani	2017	Ustvarjeno:	27. 11. 2017	Ustvarjeno:	/
			Spremenjeno:	27. 11. 2017	Spremenjeno:	27. 11. 2017
			Velikost:	41,61 MB	Velikost:	41,6 MB
			PDF različica:	1.4 (Acrobat 5.x) (PDF/A -1A)	PDF različica:	Acrobat PDF 1.4 (fmt/18)
			Aplikacija:	LuraDocument PDF Compressor		
			Producent PDF-a:	Srever 6.2.92.61 LuraDocument PDF v2.60		
			Strani:	14		

Za datoteko 01.pdf so nas zanimali podatki, ki jih lahko dobimo z orodjem *veraPDF* v obliki poročila (*Validation report*) (slika 2). Izkazalo se je, da katalog slovarja nima metapodatkov (Rule 6. 7. 2-1, slika 2).

Validation Report

File: D:\A_PREZENTACIJEVA_Prezentacije2018\Radenci\Gregor_Vanja\primeri_pdf\01.pdf
 Validation Profile: PDF/A-1B validation profile
 PDF/A compliance: **Failed**

Statistics

Version: 1.10.6
 Build Date: 2017-12-28T11:58:00+01:00
 Processing time: 00:03:41.261
 Total rules in Profile: 103
 Passed Checks: 175866
 Failed Checks: 1
 Metadata Fixes Status: Repair successful
 Completed Metadata Fixes: 3

Metadata fixes information

Fixes [Show](#)

Validation information

Rule	Status
Specification: ISO 19005-1:2005, Clause: 6.7.2, Test number: 1	
The document catalog dictionary of a conforming file shall contain the Metadata key.	Failed
1 occurrences	Hide
PDDocument	
metadata_size == 1	
root/document[0]	

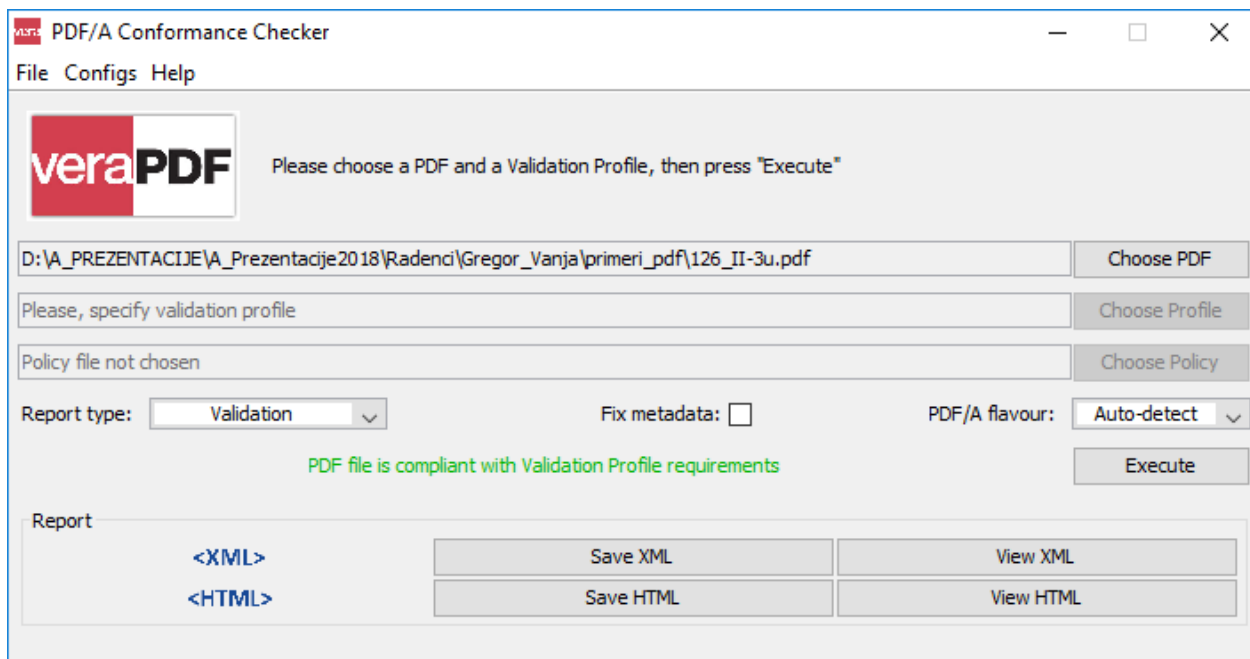
Slika 2: Poročilo validacije za datoteko 01.pdf (Rule 6.7.2-1)

Programsko orodje *veraPDF* je sicer zaznalo format PDF/A-1b, vendar je ugotovilo neskladnost s profilom tega formata, in sicer je odstopanje v pravilu 6.7.2.-1:

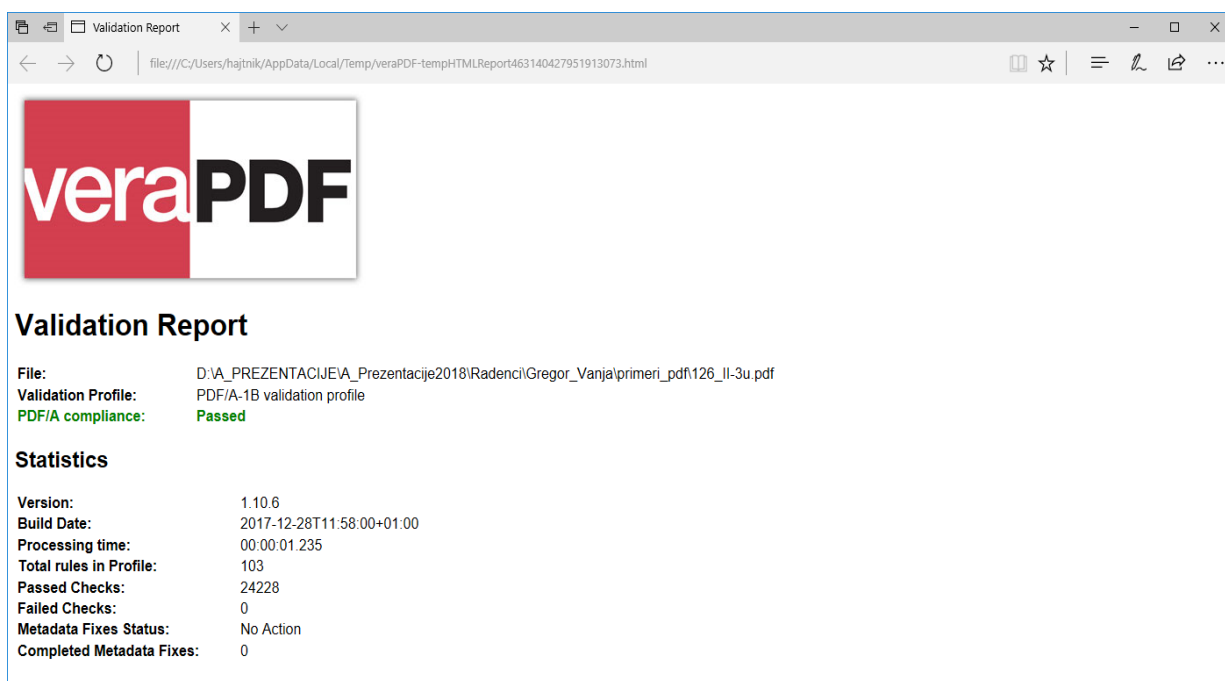
Requirement: *The document catalog dictionary of a conforming file shall contain the Metadata key.*

Error details: *The document catalog dictionary doesn't contain Metadata key.*

Vsi ostali digitalizati, za katere smo z orodjem *veraPDF* validirali format, so v formatu PDF/A-1a oz. PDF/A-1b (Tabela 2). Na slikah 3 in 4 vidimo primer uspešne validacije formata PDF/A-1b.



Slika 3: Primer uspešne validacije PDF/A



Slika 4: Primer validacijskega poročila z orodjem veraPDF

8 Zaključne ugotovitve

Skozi analizo šestih naključnih digitalizatorov Arhiva RS, ki so jih v zadnjih devetih letih digitalizirali različni ponudniki spremljevalnih storitev pretvorbe arhivskega gradiva iz analogne (tj. fizične) oblike v digitalno, se je izkazalo, da so bile vhodne uporabniške zahteve ob naročilu digitalizacije glede formata izpolnjene. Ponudniki so zagotovili izvedbo digitalizacije v zahtevanih formatih. V letu 2009 se je potrdila trditev iz uvoda prispevka, da se je digitalizacija pravzaprav izvajala »ad hoc« in po potrebi, pri čemer se ni posvečalo pretirane pozornosti dejstvu, da je bil kot izhodni format digitalizacije izbran »goli« PDF, in ne katera od danes uveljavljenih različic PDF/A, ki so obstajale že tudi v letu 2009. To je bilo odpravljeno pri digitalizatorih, nastalih po letu 2014, kar je bilo pričakovati, saj je bila takrat že pripravljena prva različica kontrolnega seznama, ki je sistemiziral postopke digitalizacije in s tem vsebinske nosilce digitalizacije »prisilil« k ukvarjanju tudi z vprašanji ustreznosti formatov. Da je bila pri tem izbrana raven b (PDF/A-1b; oz osnovna raven) formata PDF/A-1, ni idealna rešitev, a ne pomeni prevelikega odstopanja od uveljavljene prakse. PDF/-1b enako kot PDF/A-1a omogoča zvesto reprodukcijo dokumenta, hkrati pa z dodatnimi posebnostmi, kot sta na primer izbira jezika in označevanje besedila, poveča dostopnost le-tega. Kar je z vidika postopkov digitalizacije pomenljivo, pa je dejstvo, da je bila programska validacija formatov izvedena šele 1, 2, 4 oz. 9 let po naročilu digitalizacije; in to po zaključku postopka, prevzemu gradiva in končno tudi plačilu opravljene storitve. Čeprav je bila zahteva po kontroli kakovosti v postopkih digitalizacije zakonsko določena že v preteklosti, je arhivi, roko na srce, v tej razsežnosti nismo izvajali. Z napredkom tehnologij so ti postopki postali dostopnejši in predvsem izvedeni na splošno dostopni programske opreme. Nadalje predstavljen primer kaže na nujnost izvajanja arhivske zakonodaje. V tretjem odstavku 13. člena UVDAG je zakonodajalec »/.../ *samodejno oziroma ročno kontrolo pravilne izvedbe digitalizacije z namenom odprave napak oziroma odstopanj*« opredelil kot nujno, da se zagotovi avtentična pretvorba arhivskega gradiva v postopkih digitalizacije (UVDAG, 13. člen). Arhivi jo moramo pričeti izvajati, in to ne le ob koncu digitalizacije, ampak že v fazi poskusne digitalizacije, v postopkih naključnih kontrol med procesom digitalizacije kot – še posebno – na širokem vzorcu ob njenem zaključku.

SUMMARY

SUITABILITY ANALYSIS OF FORMATS FOR LONG-TERM PRESERVATION AND METADATA IN THE DIGITIZATION OF ARCHIVAL CULTURAL HERITAGE

Gregor JENUŠ, Ph. D.

Archives of the Republic of Slovenia,
Ljubljana, Slovenia
gregor.jenus@gov.si

Anja PAULIČ

Archives of the Republic of Slovenia,
Ljubljana, Slovenia
anja.paulic@gov.si

Tatjana HAJTNIK, Ph. D.

Archives of the Republic of Slovenia,
Ljubljana, Slovenia
tatjana.hajtnik@gov.si

Vanja PFAJFAR

Archives of the Republic of Slovenia,
Ljubljana, Slovenia
vanja.pfajfar@gov.si

Digitization of archival records is one of the processes for a long-term preservation of archival cultural heritage. The conversion of archival records from their analog (physical) form into a digital format serves as a method to protect the original archival records from material damage caused by frequent, or even more important, improper use. At the same time, digitization of archival cultural heritage allows for greater dissemination of archival records, since the records become more accessible in the existing reading rooms of the Slovenian archival network, as well as in the virtual archival reading room, which is currently under development within the e-ARH.si project.

In order to standardize the processes of digitization in recent years, the Slovenian archival network introduced special forms (checklists). They include information on digitization formats and associated metadata and are adapted to the specific needs of a variety of forms of archival records and designed to master the workflow of digitization:

- *preparation and inventory of archival records;*
- *definition of the digitization method;*
- *instructions for the services provider regarding handling of archival records;*
- *instructions for the services provider regarding the methods of storing before, during and after digitization, taking into account the principles of archival records preservation.*

When the outcome of digitization is an archival copy (AIP, Archival Information Package) of digitized archival records, the Slovenian archives use the PDF and/or TIFF format, while in cases of creating user copies (DIP, Dissemination Information Package) the JPEG format is being used. The development of formats, due to the needs of different users, can cause several variations within one format. Therefore, the format cannot be precisely defined only by its extension (for example .pdf). When choosing to use a format (such as PDF/A), which is suitable for long-term preservation, we must choose the format that is most in line with modern principles of digital archival preservation or other user requirements (such as accessibility).

The authors' examined already digitized archival records and tested the conformity of formats for long-term preservation, examined the relevance and integrity of various forms of metadata with the goal to optimize and complete the workflow of digitization and quality control already on the level of test digitization.

Viri in literatura

- Accessibility What PDF/A-1a Really Means. (b. d.)** Pridobljeno 28. 3. 2018 s spletne strani: <https://www.pdfa.org/accessibility-%C2%96-what-pdf-a-1a-really-means/>.
- »Arhiv RS objavil tri izvornike: Dokument SDS je sestavljen iz delov različnih dokumentom«. V: Dnevnik, 22. 2. 2011. Pridobljeno 28. 3. 2018 s spletne strani: <https://www.dnevnik.si/1042426174>.
- »Kdor ima arhiv, ima moč«. V: Dnevnik, 19. 11. 2011. Pridobljeno 28. 3. 2018 s spletne strani: <https://www.dnevnik.si/1042425331>.
- Berkeley Software Distribution.** Pridobljeno 30. 3. 2018 s spletne strani: <http://www.linfo.org/bsdlicense.html>.
- Corrado M. in Moulaison S. H. (2017).** *Digital Preservation for Libraries, Archives and Museums. 2nd Edition.* Lanham, Maryland: Rowman & Littlefield.
- Digital Preservation Handbook. (2015).** 2nd Edition, <http://handbook.dpconline.org/>.
- Digital Preservation Coalition.** Pridobljeno 28. 3. 2018 s spletne strani: <https://www.dpconline.org/handbook/technical-solutions-and-tools/file-formats-and-standards>.
- Digitalna Slovenija 2020 – Strategija razvoja informacijske družbe do leta 2020 (2016).** Pridobljeno 28. 3. 2018 s spletne strani: http://www.mju.gov.si/fileadmin/mju.gov.si/pageuploads/DID/Informacijska_druzba/DSI_2020.pdf.
- Dobernik, M. (2002).** Reproduciranje franciscejskega katastra. V: *Arhivi. Glasilo Arhivskega društva in arhivov Slovenije*, 25 (2), str. 65–76. Ljubljana: Arhivsko društvo Slovenije.
- Enotne tehnološke zahteve I., II., III. del. (2013).** Pridobljeno 28.3.2018 s spletne strani: http://www.arhiv.gov.si/si/zakonodaja_in_dokumenti/ .
- Evropska digitalna agenda (2010).** COM(2010) 245 konč./2, 26. 8. 2010. Pridobljeno 28. 3. 2018 s spletne strani: https://europa.eu/european-union/file/1529/download_sl?token=x0E0KD8s.
- File format.** Pridobljeno 28. 3. 2018 s spletne strani: https://en.wikipedia.org/wiki/File_format.
- GeoTIFF, Revision 1.0 (2011).** Pridobljeno 28. 3. 2018 s spletne strani: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000279.shtml>.
- Graphic File Formats (2008).** Pridobljeno 28. 3. 2018 s spletne strani: <http://www.nationalarchives.gov.uk/documents/information-management/graphic-file-formats.pdf>.
- Best Image Formats for Archival Photography (2016).** Pridobljeno 28. 3. 2018 s spletne strani: <https://digitalarch.org/blog/2017/4/7/best-image-formats-for-archival-photography>.
- Hajtnik, T. (2016).** *Celovit pristop k pretvorbi elektronskih dokumentov v obliko za dolgoročno hrambo.* Doktorska disertacija. Maribor: Fakulteta za elektrotehniko, računalništvo in informatiko.
- Hajtnik, T. et al. (2009).** *Arhiviranje, hramba in upravljanje dokumentov. Navodila in praktični primeri pravilne hrambe vseh vrst dokumentarnega gradiva za podjetja in ustanove.* Maribor: Forum Medija.

- International Standard for Archival Description (ISAD(G)).* (2000). 2nd Edition. Ottawa: International Council on Archives. Pridobljeno 28.3.2018 s spletne strani: <https://www.ica.org/en/isadg-general-international-standard-archival-description-second-edition>.
- JPEG Image Encoding Family.** (2012). Pridobljeno 28. 3. 2018 s spletne strani: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000017.shtml>.
- Kontrolni seznam za digitalizacijo arhivskega gradiva.** (2016). Verzija 1.6. Ljubljana: Arhiv Republike Slovenije.
- Nacionalni program za kulturo 2008-2014.** (2008). (Resolucija, Ur. l. RS 35/08 in 95/10). Pridobljeno 28. 3. 2018 s spletne strani: <https://www.uradni-list.si/glasilo-uradni-list-rs/vsebina/2008-01-1428?sop=2008-01-1428>.
- Nacionalni program za kulturo 2004-2007.** (2004). (Resolucija, Ur. l. RS 28/04). Pridobljeno 28. 3. 2018 s spletne strani: <https://www.uradni-list.si/glasilo-uradni-list-rs/vsebina/2004-01-1205?sop=2004-01-1205>.
- PDF/A. (b. d.).** Pridobljeno 28. 3. 2018 s spletne strani: <https://en.wikipedia.org/wiki/PDF/A>.
- PDF/A-1, PDF for Long-term Preservation, Use of PDF 1.4.** (2010). Pridobljeno 28. 3. 2018 s spletne strani: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000125.shtml#useful>.
- Pfajfar, V. (2018).** *Digitalizacija arhivskega gradiva v arhivih. Metodologija in standardizacija postopkov.* Magistrsko delo. Maribor: Alma Mater Europea.
- PREFORMA.** Pridobljeno 30. 3. 2018 s spletne strani: <http://www.preforma-project.eu/>.
- PRONOM.** Pridobljeno 30. 3. 2018 s spletne strani: <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>.
- Strategija in izvedbeni načrt razvoja slovenskega elektronskega arhiva za obdobje 2010 – 2015.** (2010). Pridobljeno 28. 3. 2018 s spletne strani: http://www.arhiv.gov.si/fileadmin/arhiv.gov.si/pageuploads/zakonodaja/Predlog_str_e-ARH.si_1.5.pdf.
- Strategija in izvedbeni načrt razvoja slovenskega elektronskega arhiva eARH.si za obdobje 2016-2020.** (2016). Pridobljeno 28. 3. 2018 s spletne strani: http://www.arhiv.gov.si/fileadmin/arhiv.gov.si/pageuploads/zakonodaja/Strategija_e-ARH.si/Strategija_e-ARH_si_2016-2020_1.0.pdf.
- Strategija za enotni digitalni trg.** (2015). Pridobljeno 28. 3. 2018 s spletne strani: http://europa.eu/rapid/press-release_IP-15-4653_sl.htm.
- Thomas, S. in Martin, J. (2006).** Using the papers of contemporary British politicians as a testbed for the preservation of digital personal archives. V: *Journal of the Society of Archivists* 27 (1), str. 29–56.
- Uredba o varstvu dokumentarnega in arhivskega gradiva.** (2006). Uradni list Republike Slovenije, št. 86.
- Uredba o varstvu dokumentarnega in arhivskega gradiva.** (2017). Uradni list Republike Slovenije, št. 42.
- Zakon o varstvu dokumentarnega in arhivskega gradiva ter arhivih.** (2006). Uradni list Republike Slovenije, št. 30.
- Zakon o spremembah in dopolnitvah Zakona o varstvu dokumentarnega in arhivskega gradiva ter arhivih.** (2014). Uradni list Republike Slovenije, št. 51.