

Prejeto / Received: 26. 03. 2018
<https://doi.org/10.54356/MA/2018/ZZRB5571>

1.03 Kratski znanstveni članek

1.03 Short Scientific Article

TEHNIČNI VIDIKI DOLGOROČNEGA OHRANJANJA ELEKTRONSKE POŠTE

Klavdija Krivec

Arhiv Republike Slovenije, Ljubljana, Slovenija
klavdija.krivec@gov.si

dr. Tatjana HAJTNIK

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

Izvleček:

V okviru nadaljnjega razvoja projekta e-ARH.si, katerega podlago predstavljata Strategija in izvedbeni načrt razvoja slovenskega elektronskega arhiva za obdobje 2016–2020, se pripravljajo rešitve za dolgoročno ohranjanje in arhiviranje e-pošte. V prispevku je obravnavana problematika formatov dolgoročne hrambe za e-pošto, možnosti pretvorb iz izvornih v arhivske formate ter njihova validacija. Predstavljene so smernice iskanja rešitev za hrambo e-pošte najpogostejše zastopanih namiznih poštnih odjemalcev naših ustvarjalcev, IBM Lotus Notes in Microsoft Outlook. Na podlagi ugotovitev so zasnovani možni scenariji izvoza in pretvorbe e-pošte, ki predstavljajo izhodišča za sprejemanje nadaljnjih odločitev pri izbiri ustreznih tehničnih rešitev, namenjenih dolgoročnemu ohranjanju e-pošte pri ustvarjalcih in v arhivskih institucijah.

Ključne besede:

projekt e-ARH.si: ESS 2016–2020, dolgoročna hramba e-pošte, PDF/A format, pretvorba formatov e-pošte

Abstract:

Technical Aspects of Long-Term E-Mail Preservation

As a part of further development of the e-ARH.si project, based on the Strategy and Implementation Plan for the Development of the Slovenian Electronic Archives for the period 2016 - 2020, the preparation of solutions for the long-term email preservation is in progress. Article summarizes e-mail format issues, the possibility of transformation into archival formats and their validation. According to the survey, the authors focused on an email solution for the most frequently used email applications of our creators, IBM Lotus Notes and Microsoft Outlook. Possible scenarios of exports and conversions were designed for the purpose of making further decisions in the selection of appropriate technical solutions for the long-term email preservation at creators and in archival institutions.

Key words:

project e-ARH.si: ESS 2016 - 2020, long-term email preservation, PDF/A format, email formats conversion

1. Razvoj slovenskega elektronskega arhiva e-ARH.si

Slovenski javni arhivi so zadolženi za varovanje dokumentacije, ki ima poseben pomen za znanost in kulturo ter pravno varnost posameznikov in države, tudi v elektronski obliki. Množična uporaba informacijske tehnologije v sodobni družbi na vseh področjih in na vseh ravneh njenega delovanja vodi v množico pojavnih oblik dokumentarnega in arhivskega gradiva, ki se uresničuje v fizičnih in elektronskih oblikah. Ocenjujemo, da bo do leta 2020 velik del arhivskega gradiva hranjen samo še v elektronski obliki (Hajtnik, 2011).

V Sloveniji v ta namen vzpostavljamo slovenski javni elektronski arhiv e-ARH.si, ki predstavlja kompleksno okolje, namenjeno elektronskemu arhiviranju. Zajema postopke prevzemanja elektronskega arhivskega gradiva (v nadaljevanju eAG) v pristojne arhive, njegovo strokovno obdelavo, izvedbo dolgoročne hrambe v skladu z veljavno zakonodajo ter omogočanje učinkovitega upravljanja in uporabo tega gradiva. Temelj za izgradnjo je predstavljala v letu 2010 s strani Vlade RS potrjena *Strategija razvoja slovenskega javnega elektronskega arhiva e-ARH.si* in njen izvedbeni načrt. V okviru strategije e-ARH.si 2010–2015 je bila razvita prva rešitev elektronskega arhiviranja, s katero lahko Arhiv RS od ustvarjalcev prevzema eAG. Nadaljnji razvoj rešitev elektronskega arhiviranja predstavlja *Strategija in izvedbeni načrt razvoja slovenskega elektronskega arhiva 2016–2020*, potrjena marca 2016 s strani Vlade Republike Slovenije. Za njeno uresničevanje je Arhiv RS jeseni leta 2016 uspel pridobiti finančna sredstva iz Operativnega programa za izvajanje Evropske kohezijske politike v obdobju 2014–2020. S tem so bili vzpostavljeni vsi pogoji za vzpostavitev projekta nadaljnjega razvoja e-ARH.si, poimenovanega »Projekt e-ARH.si: ESS 2016–2020«.

Namen projekta je skrb za eAG pri ustvarjalcih in njegov prevzem v e-ARH.si, skrb za njegovo nadaljnje dolgoročno ohranjanje ter uporabo in povečanje njegove dostopnosti. Posebna pozornost je namenjena dostopu in nadaljnji uporabi arhivskega gradiva vsem bodočim uporabnikom, ne glede na njihove zmožnosti, tudi starejšim in osebam s posebnimi potrebami. V okviru projekta e-ARH.si: ESS 2016–2020 se posvečamo nadaljnjemu razvoju sistema e-ARH.si ter se pri tem osredotočamo na dolgoročno ohranjanje novih vrst eAG, kot so elektronska pošta (v nadaljevanju e-pošta), avdio- in videovsebine, spletne strani, družbena omrežja, prostorski podatki in 3D-modeli.

Slovenski elektronski arhiv e-ARH.si temelji na mednarodnem standardu ISO 14721:2003 oziroma referenčnem modelu OAIS (Slika 1). Model OAIS razlikuje tri oblike informacijskih paketov: sprejemni informacijski paket (SIP), arhivski informacijski paket (AIP) in dostopni informacijski paket (DIP). Ustvarjalec izroči eAG pristojnemu arhivu kot zaokroženo celoto v SIP-u s pomočjo programskega orodja Urejevalnik SIP (e-ARH.siP). Za dolgoročno ohranjanje eAG je vzpostavljen arhivski informacijski sistem (AIS), ki omogoča hrambo eAG v obliki AIP. S pomočjo Pregledovalnika DIP (e-ARH.siDIP) je končnim uporabnikom omogočen dostop do prevzetega eAG. Ta vrne dostopni informacijski paket, sestavljen iz vsebine enega ali več AIP, kot odgovor na poizvedbo in naročilo v arhivskem informacijskem sistemu.

2. Elektronska pošta

Prvo e-poštno sporočilo (v nadaljevanju e-sporočilo), ki je bilo poslano leta 1971, je popolnoma spremenilo način komuniciranja med posamezniki in skupinami širom sveta. Danes izmenjava e-pošte predstavlja bistveni del spletnih komunikacij, te pridobivajo pomen in veljavo tudi pri ustvarjalcih javnega arhivskega gradiva. Dolgoročno ohranjanje e-pošte v arhivu vključuje njen prevzem od ustvarjalcev, nadaljnje dolgoročno ohranjanje po načelih varne dolgoročne hrambe (dostopnost, uporabnost, celovitost, avtentičnost, trajnost), kot jih določa tudi Zakon o varstvu dokumentarnega in arhivskega gradiva ter arhivih (ZVDAGA), in zagotavljanje nadaljnje dostopnosti do tega gradiva vključno z možnostjo njegove ponovne uporabe za bodoče uporabnike.

V prispevku so predstavljeni do sedaj raziskani tehnični vidiki arhiviranja e-pošte, ki predstavljajo začetni korak za vzpostavitev stabilnega in zaupanja vrednega slovenskega elektronskega arhiva e-ARH.si do leta 2020 tudi za to vrsto eAG. V času projekta e-ARH.si: ESS 2016–2020 bomo poiskali programska orodja in prilagodili že obstoječa orodja za ustvarjalce, arhiviste in uporabnike arhivskega gradiva, ki bodo omogočila pripravo in prevzem e-pošte v sistem e-ARH.si in njegovo nadaljnjo uporabo.

3. Tehnični izzivi pri dolgoročni hrambi e-pošte

Urejevalnik SIP, namenjen izročanju eAG, zaenkrat podpira sprejem gradiva v obliki nestrukturiranih zapisov, podatkovnih zbirk v obliki relacijskih baz ter gradiva iz informacijskih sistemov za upravljanje z dokumenti. V času projekta e-ARH.si do leta 2020 želimo omogočiti ustvarjalcem tudi izročanje e-pošte. Na podlagi do sedaj raziskanih možnosti arhiviranja e-pošte bo ta v Urejevalniku SIP obravnavana kot nestrukturirani zapisi. Datoteke e-pošte bodo kot vsebina ene ali več map uvožene v sam proces gradnje paketa, nato sledi urejanje metapodatkov za posamezno opisno enoto ter izvoz opisnih enot s pripadajočim naborom metapodatkov v SIP. Izjemnega pomena je kvalitetna predpriprava datotek gradiva e-pošte za prevzem z Urejevalnikom SIP, kjer je na eni strani pomemben vsebinski vidik, ki je domena arhivske stroke, in na drugi strani tehnični vidik, obravnavan v prispevku.

S tehnične perspektive smo se v začetnih korakih usmerili v iskanje odgovorov na vprašanja, kateri so ustrezni formati za dolgoročno hrambo e-pošte v arhivih po svetu, kateri formati e-pošte so v uporabi pri naših ustvarjalcih, kako lahko pretvorimo izvirne formate e-pošte v formate za dolgoročno hrambo in kako ustrezno izvedemo validacijo pretvorjenih datotek. S tehničnega vidika se pri arhivskem gradivu v obliki e-pošte spoprijemamo z izzivi vključevanja te vrste gradiva v celoten zgoraj opisani proces, razširjen s postopki prevzemanja e-pošte v pristojne arhive, njeno strokovno obdelavo, dolgoročno hrambo v skladu z veljavno zakonodajo ter omogočanjem nadaljnjega učinkovitega upravljanja in uporabe.

4. Formati dolgoročne hrambe za e-pošto

Velik izziv, za nas in strokovnjake v svetu, je izbrati ustrezen format za dolgoročno hrambo e-sporočil (Hajtnik, 2012). Izvirna e-sporočila se hranijo v različnih formatih datotek. Trenutno uveljavljena strategija dolgoročnega hranjenja e-dokumentov je pretvorba iz izvirne oblike v obliko zapisa, ki je primerna za dolgoročno hrambo. Pri nas je določena kot obvezna strategija tudi z ZVDAGA (11. člen). Medtem ko lahko pretvorba

formatov datotek ščiti podatke pred izgubo, pa sam postopek pretvorbe prinaša določena tveganja, zato ga je treba skrbno načrtovati in izvajati.

E-sporočilo je sestavljeno iz glave (angl. *header*) in telesa (angl. *body*). Glava vključuje prejemnikov e-naslov in opsijsko ime prejemnika ali prejemnikov, pošiljateljev e-naslov in opsijsko ime, čas in datum sporočila, vključno s časovnim pasom ter predmet sporočila. Telo e-sporočila lahko vsebuje besedilo sporočila in pripete datoteke oziroma priponke (angl. *attachments*), ki jim je pri izbiri formata za arhiviranje e-pošte namenjena posebna pozornost. Preko e-pošte lahko pošljemo poljubne vrste datotek vključno z video posnetki, glasbo, slikami in dokumenti, odpremo pa jih s programi, ki niso poštni odjemalec. Arhiviranje e-pošte zajema tudi arhiviranje priponk poljubnih formatov, čemur smo se še posebej posvetili pri izbiri formatov za arhiviranje e-pošte.

V svetu se kot primeren format za hrambo e-sporočil sprejema tudi format PDF/A, določen v standardu ISO 19005. PDF/A je usmerjen v zanesljivo dolgoročno shranjevanje e-dokumentov z besedilom, rastrskimi slikami in vektorsko grafiko ter z njimi povezanimi metapodatki. Format PDF/A si prizadeva zagotoviti dosledno in robustno podmnožico formata PDF¹, ki se lahko reproducira tudi po daljšem časovnem obdobju hrambe in se uporablja za zanesljivo izmenjavo podatkov v podjetjih in vladnih okoljih (Prom, 2011). V nadaljevanju si bomo ogledali njegove podtipе PDF/A-1, PDF/A-2 in PDF/A-3.

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-ja: zahteva strukturne informacije in zanesljivo besedilno semantiko Unicode², 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 PDF).

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, PDF/A-2 ponuja 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 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.

PDF/A-3 je bil objavljen 2012 kot ISO 19005-3 in podpira skladne ravni a, b in u. PDF/A-3 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.

¹ PDF (angl. *Portable Document Format*) je format za izmenjavo e-dokumentov, definiran v ISO 32000-1:2008. Je odprt standard za izmenjavo e-dokumentov, ki je bil ustvarjen z namenom neodvisnega prikazovanja dokumenta od platforme (programska oprema, strojna oprema, operacijski sistem).

² 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.

5. Pretvorba izvirnih formatov e-pošte v arhivski format ter validacija pretvorjenih datotek

Naši ustvarjalci javnega arhivskega gradiva uporabljajo različne odjemalce e-pošte, ki jim omogočajo izvoz v različne formate. S pomočjo pretvornikov e-sporočil je mogoče izvesti pretvorbe med različnimi standardnimi formati (npr. Lotus Notesov format .nsf v Outlookov format .pst) ter pretvorbe iz formatov e-pošte v enostavnejše formate (.xml). V produkcijskem okolju javne uprave se najpogosteje pojavljata e-odjemalca Microsoft Outlook in Lotus Notes ter njima pripadajoči formati arhiviranja: Outlookov format .pst ter Lotus Notesov format .nsf. Izbira ustreznih programskih orodij za pretvorbo je izjemnega pomena, saj so na voljo *lastniška*³, *brezplačna*⁴ ali *odprtokodna*⁵ orodja z različnimi funkcionalnostmi in cenovno dostopnostjo. Na podlagi pregleda dobrih praks v svetu smo se usmerili v iskanje in testiranje različnih pretvornikov e-pošte iz izvirnih v ustreznejše splošne formate, primerne za dolgoročno hrambo, še posebej v format PDF/A.

Pretvorbi iz izvirnega v željeni format sledi validacija oziroma kontrola uspešnosti pretvorbe. Standardi, na katerih temelji format PDF/A, niso vedno enoznačni. Posledično se na tržišču za validacijo formata PDF/A pojavljajo programska orodja, ki so v lasti posameznih komercialnih podjetij ter izvajajo validacijo pretvorjenih datotek na podlagi različnih kriterijev, kar vodi v nekonsistentne rezultate ter neučinkovito določevanje skladnosti s formatom PDF/A. Potreba po zanesljivem odprtokodnem programskem orodju ter interoperabilnosti je sprožila financiranje razvoja ustreznega programskega orodja s strani projekta Evropske unije PREFORMA – PREservation FORMAts, ki ga podpira združenje predstavnikov digitalne hrambe (Open Preservation Foundation, Digital Preservation Coalition, KEEP Solutions) in vodilnih članov skupnosti razvijalcev programske opreme PDF (PDF Association in Dual Lab). Glavni namen projekta naslavlja implementacijo standardiziranih formatov datotek za dolgoročno hrambo ter s tem želi omogočiti arhivskim institucijam nadzor nad sprejemanjem in upravljanjem datotek za shranjevanje v digitalne repozitorije. Za validacijo formata PDF/A je bilo razvito odprtokodno programsko orodje, ki preveri skladnost datotek s standardnimi specifikacijami, skladnost datotek z merili oz. kriteriji institucije, izdela poročilo ter izvaja preproste metapodatkovne popravke. Orodje omogoča validacijo osmih profilov PDF/A, to so PDF/A-1b, -1a, -2b, -2u, -2a, -3b, -3u, -3a. 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 (angl. Application programming interface, Java API) za implementacijo funkcionalnosti orodja v lastne programske rešitve. Trenutno je na voljo zadnja različica veraPDF 1.10, izdana v novembru 2017. Orodje je izdano pod licencama GPLv3⁶ in MPLv2⁷.

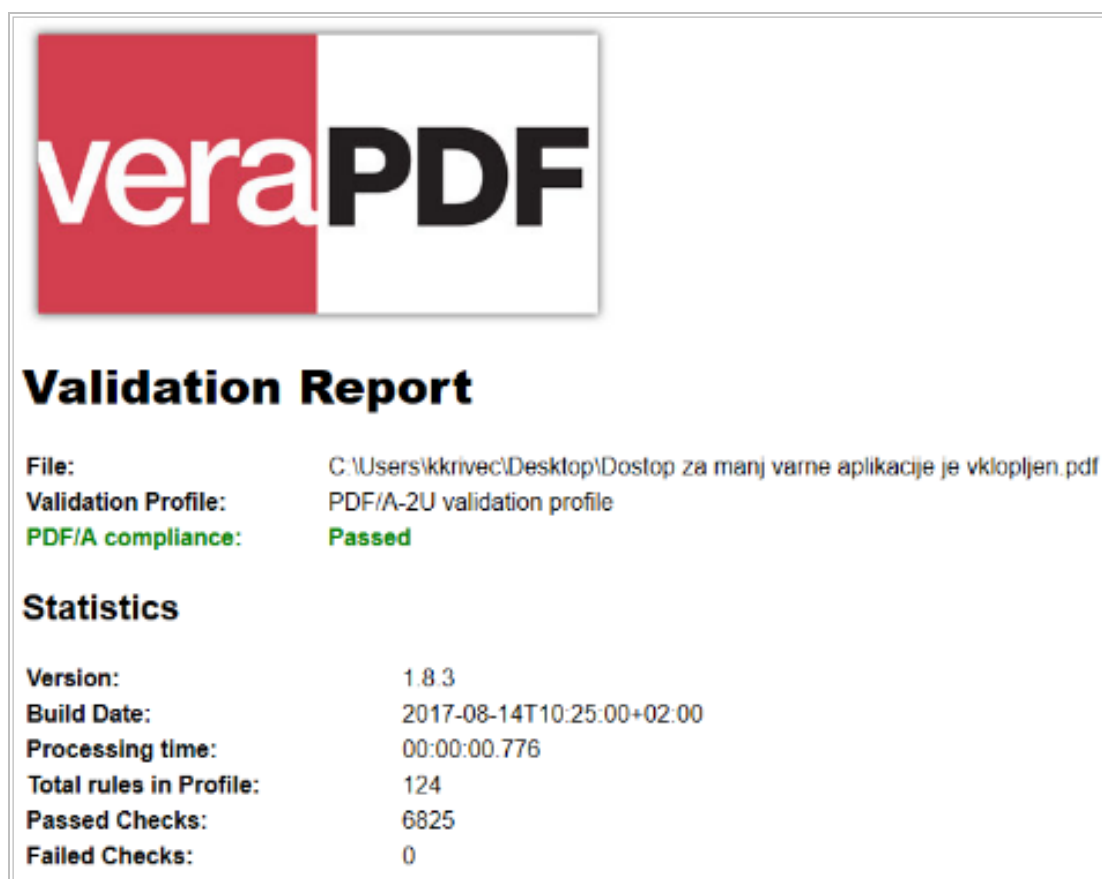
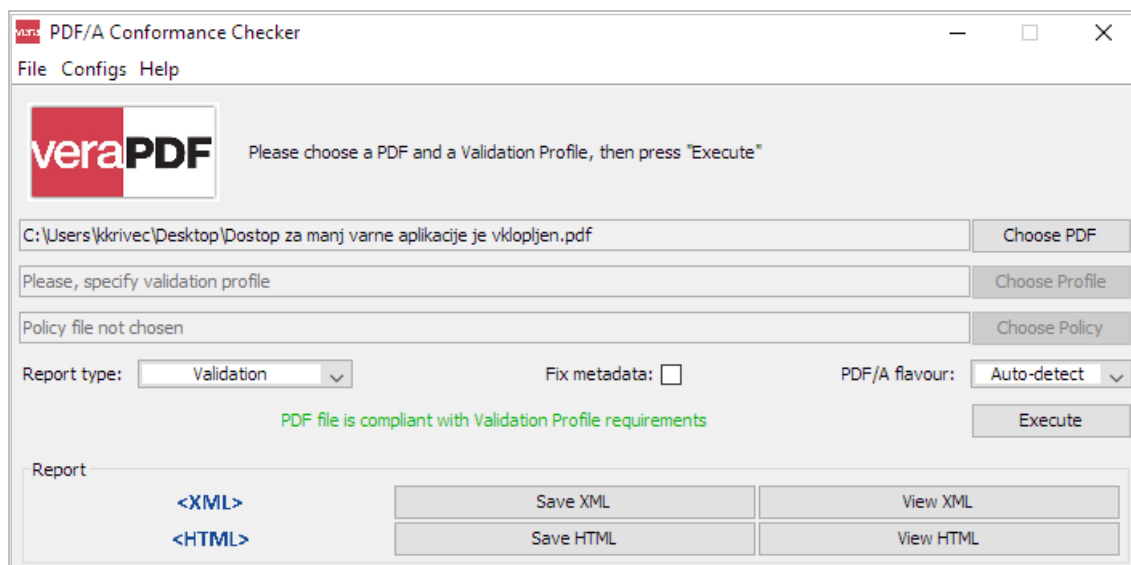
³ *Lastniška programska oprema (angl. Proprietary software) je programska oprema, za katero izdajatelj ali druga oseba ohrani pravice intelektualne lastnine, običajno avtorske pravice za izvirno kodo, včasih tudi patentne pravice.*

⁴ *Brezplačna programska oprema (angl. Freeware) je programska oprema, ki je na voljo na spletu, brez plačila ali za prostovoljno plačilo, največkrat brez programske kode.*

⁵ *Odprtokodna programska oprema (angl. Open-source software, OSS) je programska oprema, za katero je izvirna koda izdana pod zaščitno licenco, ki ustreza kriterijem odprtokodne definicije. Ta dovoljuje vsakršno uporabo in spreminjanje programske opreme in njeno razširjanje.*

⁶ *GNU General public license, dostopno na <https://www.gnu.org/licenses/gpl-3.0.en.html>.*

⁷ *Mozilla Public License Version 2.0, dostopno na <https://www.mozilla.org/en-US/MPL/>.*



Slika 1: Primer validacije s programsko opremo veraPDF

6. Analiza uporabe e-pošte pri ustvarjalcih javnega arhivskega gradiva

Novembra 2016 je bila med ustvarjalci javnega arhivskega gradiva izvedena anketa, s katero smo med drugim pridobili tudi osnovne informacije s področja uporabe e-pošte pri njih.

Anketni vprašalnik je bil zasnovan z namenom omogočiti vpogled v arhivsko gradivo v obliki e-pošte, ki obstaja pri ustvarjalcih javnega arhivskega gradiva. Na vzorcu 52 ustvarjalcev, ki so podali odgovore na anketna vprašanja, smo pridobili osnovne informacije o rabi e-pošte ter na podlagi analize zbranih podatkov zasnovali smernice nadaljnjega raziskovanja. Zanimalo nas je predvsem, v kolikšni meri pri ustvarjalcih nastaja arhivsko gradivo v obliki e-pošte in kako je omogočen dostop do e-sporočil.

Analiza rezultatov zbranih podatkov na našem vzorcu ustvarjalcev je pokazala sledeče:

- arhivsko gradivo v obliki e-pošte nastaja pri vseh anketiranih ustvarjalcih,
- dostop do e-pošte je omogočen s pomočjo namiznih poštnih odjemalcev⁸ ali spletnih vmesnikov⁹,
- najpogostejša namizna poštna odjemalca sta IBM Lotus Notes in Microsoft Outlook, drugi odjemalci so v manjšini.

Oglejmo si razlike med namiznimi poštnimi odjemalci in spletno pošto. Namizni poštni odjemalci so aplikacije oziroma programska orodja, ki nam omogočajo upravljanje z e-pošto. Preko teh programov poteka sestavljanje, pošiljanje in prejemanje e-sporočil, ki se shranjujejo na uporabnikovem lokalnem računalniku. Predhodno je treba v poštnem odjemalcu le nastaviti e-poštni račun, kar vključuje vnos temeljnih nastavitev, kot so uporabniško ime, poštni naslov in geslo, nastavitve strežnikov za dohodno (POP3¹⁰ ali IMAP¹¹) in odhodno pošto (SMTP¹²) ter vrsta šifrirane povezave. Vzporedno s poštnimi odjemalci se je začel razvijati dostop do spletne pošte (Arnes, Gmail, Yahoo! Mail) preko spletnega vmesnika s pomočjo spletnih brskalnikov (npr. Internet Explorer, Mozilla Firefox, Google Chrome). Ustvarjalci upravljajo z e-pošto tudi preko spletnih vmesnikov za spletno pošto. E-sporočila so shranjena na spletnem strežniku, pošiljanje in prejemanje sporočil pa poteka preko spletnega brskalnika na spletni strani e-poštne storitve. Glavna prednost spletne pošte je zmožnost pošiljanja in prejemanja e-pošte kjerkoli iz spletnega brskalnika, njena glavna pomanjkljivost pa je potreba po povezavi z internetom, medtem ko jo uporabljamo.

⁸ Namizni poštni odjemalec (angl. *Email Client*) oz. e-poštni odjemalec je namizna aplikacija, ki omogoča konfiguriranje enega ali več e-poštnih naslovov za prejemanje, branje, sestavljanje in pošiljanje e-sporočil. E-poštni odjemalec je znan tudi kot uporabniški agent za pošto (angl. *Mail Transfer Agent*, MUA).

⁹ Spletna pošta z dostopom preko spletnega vmesnika se izvaja kot spletna aplikacija na spletnem strežniku.

¹⁰ POP3 (angl. *Post Office Protocol version 3*) je bil prvič definiran v standardu RFC 1081 leta 1988 (<https://tools.ietf.org/html/rfc1081>). Je internetni protokol na aplikacijskem sloju, ki je namenjen pridobivanju elektronske pošte iz strežnika na odjemalniški računalnik preko TCP/IP-povezave.

¹¹ IMAP (angl. *Internet Message Access Protocol*) je bil prvič definiran v standardu RFC 3501 leta 2003 (<https://tools.ietf.org/html/rfc3501>). Je internetni protokol na aplikacijskem sloju preko TCP/IP-povezave, ki se uporablja kot alternativa protokolu POP3.

¹² SMTP (angl. *Simple Mail Transfer Protocol*) je bil prvič definiran v standardu RFC 821 leta 1982 (<https://tools.ietf.org/html/rfc821>). Je internetni protokol na aplikacijskem sloju preko TCP/IP-povezave, namenjen pošiljanju sporočil.

Široko razširjena uporaba e-pošte pred nas postavlja izzive zagotavljanja dolgoročne hrambe tudi za to vrsto arhivskega gradiva. Zbrani rezultati so predstavljali izhodišča za proučevanje tehničnih vidikov zagotavljanja dolgoročne hrambe e-pošte. Naši ustvarjalci v največji meri dostopajo do e-pošte s poštnima odjemalcema Microsoft Outlook in IBM Lotus Notes, raba ostalih načinov dostopa in upravljanja z e-pošto preko drugih poštnih odjemalcev ali preko spletnega vmesnika je manj razširjena. V nadaljnjem raziskovanju smo se osredotočili v proučevanje tehničnih možnosti zagotavljanja dolgoročne hrambe in arhiviranja e-pošte za omenjena namizna poštna odjemalca, za katera smo zasnovali tehnični scenarij izvoza in pretvorbe e-pošte. Osnovni scenarij je mogoče razširiti na način, da se podpre obstoječe alternative dostopov do e-pošte, ki se pojavljajo pri naših ustvarjalcih. Spletno pošto je mogoče s spletnega strežnika uvoziti v željeni poštni odjemalec (Microsoft Outlook ali Lotus Notes), predhodno moramo v odjemalcu le opredeliti nastavitve poštnega računa. Za ustvarjalce, ki uporabljajo druge e-poštne odjemalce, je možno izvoziti e-pošto v željeni poštni odjemalec s pomočjo funkcionalnosti izvoza v izvirnem poštnem odjemalcu in uvoza v željenem poštnem odjemalcu ali preko tretjega programa, ki poskrbi za pretvorbo elektronskih sporočil.

7. Zasnovani scenariji izvoza in pretvorbe za e-pošto

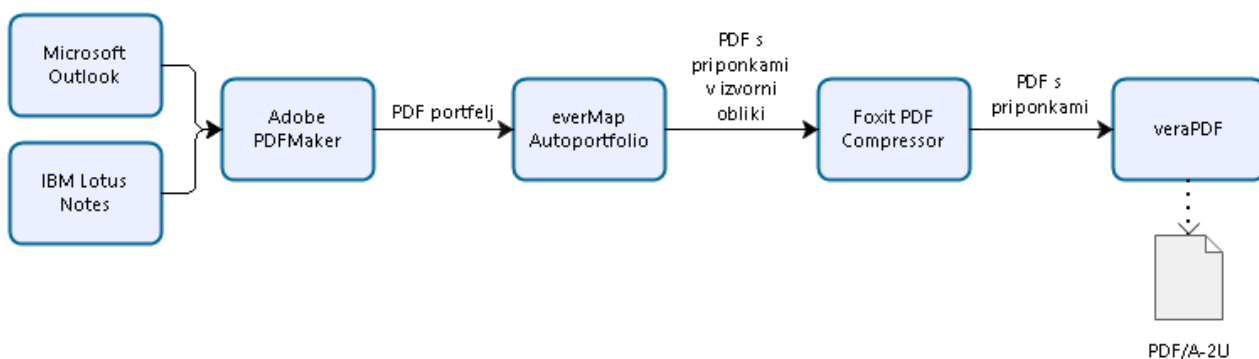
Iskanje najustreznejših programov za pretvorbo formatov je ena ključnih nalog pri arhiviranju e-pošte. Glede na zasnovane kriterije je bila izvedena izbira najustreznejše programske opreme za izvedbo pretvorbe formatov e-pošte, ustvarjene s programoma Outlook in Lotus Notes, v arhivski format PDF/A. Na podlagi preverjanja različnih aktualnih pretvornikov smo oblikovali možne scenarije pretvorbe, pri čemer smo pri izbiri upoštevali različne kriterije izbora, kot so:

- dostopnost testne različice,
- enostavnost namestitve,
- preprostost in razumljivost uporabe orodja,
- pretvorba e-sporočil in priponk iz izvornih v arhivski format PDF/A,
- ohranjanje strukture direktorijev in poddirektorijev,
- možnost ugotovitve ali pretvorba robnih primerov, kot je npr. format .ZIP,
- osnovna programska in strojna oprema (operacijski sistem, programska oprema, prostor na disku, zmogljivost procesorja),
- morebitna dodatna programska in strojna oprema,
- možnost razvoja lastnih razširitev,
- cenovna dostopnost ipd.

Scenariji pretvorbe

Prvi scenarij

Pretvorbo izvedemo s pomočjo programskih orodij Adobe Acrobat PDFmaker, Autoportfolio in Foxit PDF Compressor (**Napaka! Vira sklicevanja ni bilo mogoče najti.**2). Orodje Adobe Acrobat PDFmaker je funkcija programa Acrobat, ki deluje v množici poslovnih programov, med drugim ga lahko kot programski dodatek (angl. *plug-in*) namestimo tudi v Microsoft Outlook in IBM Lotus Notes. Z dodatkom PDFMaker lahko pretvorimo več e-sporočil ali celotne mape sporočil v portfelj PDF ali v spojeno datoteko PDF. Zaradi kasnejše pretvorbe posameznega e-sporočila s pripadajočimi priponkami v format PDF/A je za nas zanimiva opcija pretvorbe v portfelj, saj so posamezna e-sporočila sicer združena v eno datoteko, iz vizualizacije datoteke s programom Adobe Reader pa je razvidna struktura pretvorjenih sporočil in map po posameznih mapah in e-sporočilih ter tudi pripadajočih priponkah. Pri pretvorbi se v pravilni hierarhiji ohrani prvotna struktura posameznih e-sporočil s pripadajočimi priponkami in nadrejenih ali podrejenih map. Rezultat izvoza je portfelj PDF, ki je zbirka PDF-dokumentov. S pomočjo programskega dodatka Autoportfolio za programsko opremo Adobe Acrobat portfelj PDF razdelimo na posamezne datoteke PDF. Datoteko PDF sestavlja osnovno e-sporočilo s pripadajočimi priponkami. Validacija datoteke PDF pokaže, da je format datoteke PDF 1.7. Format nato z orodjem Foxit PDF Compressor pretvorimo v format PDF/A. Validacija z orodjem veraPDF pokaže, da je pretvorjen format datoteke arhivski (PDF/A-2U).



Slika 2: Scenarij pretvorbe e-pošte z orodji Acrobat PDFMaker, Autoportfolio in Foxit PDF Compressor

Drugi scenarij

Pretvorba vključuje programska orodja Aid4mail, Stellar NSF to PST Converter, SysTools NSF Converter in Foxit PDF Compressor (**Napaka! Vira sklicevanja ni bilo mogoče najti.**3). Poštna odjemalca IBM Lotus Notes in Microsoft Outlook sama po sebi ponujata možnost izvoza oziroma arhiviranja. Izvoz vsebine e-pošte s priponkami ter po želji tudi izvoz podatkov o uporabniku, stikov, informacij v koledarju ipd. ustvari kopijo informacij, saj vsi podatki ostanejo v poštnem odjemalcu in so tam tudi dostopni. Ustvarjena datoteka je shranjena v našem računalniku na izbranem mestu. Ustvarjena podatkovna datoteka v Outlooku je izvožena v izvornem formatu .pst, datoteka Lotus Notes pa v izvornem formatu .nsf.

Outlook:

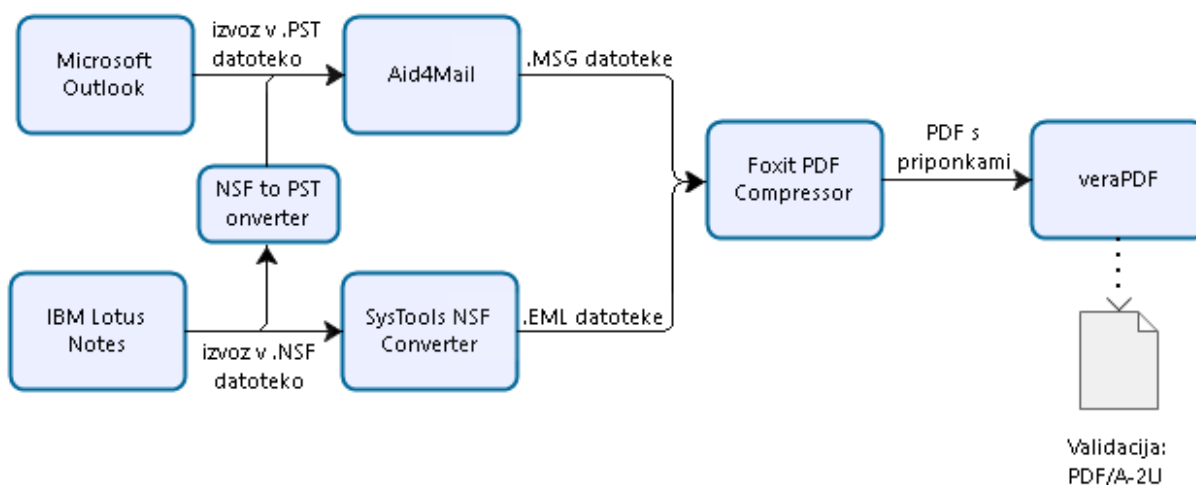
V nadaljevanju s programskim orodjem Aid4Mail pretvorimo izvozno datoteko .pst, v katero smo izvozili izbrane mape poštnega predala, v datoteke .msg posameznih e-

sporočil. Nato posamezno datoteko .msg s PDF Compressorjem pretvorimo v format PDF s priponkami.

Lotus Notes:

Možnost 1: Podobno v Lotus Notesu izvozno datoteko .nsf s pretvornikom »NSF to PST Converter« (Stellar Office) pretvorimo v datoteko .pst in nadaljujemo s procesom pretvorbe kot pri datoteki .PST.

Možnost 2: Izvozno datoteko .nsf z orodjem »SysTools NSF Converter« pretvorimo v formata .eml ali .msg, nato pa postopek nadaljujemo s pretvorbo datotek .eml ali .msg v format PDF s priponkami.



Slika 3: Scenarij pretvorbe e-pošte s programskimi orodji Aid4mail, Stellar NSF to PST Converter, SysTools NSF Converter, Foxit PDF Compressor.

Validacija z orodjem veraPDF pokaže, da je pretvorjen format datotek arhivski (PDF/A-2U).

8. Zaključek

V prispevku smo prikazali prve rezultate projekta e-ARH.si: ESS 2016–2020 na področju zagotavljanja dolgoročne e-pošte kot ene izmed tistih vrst arhivskega gradiva, ki nastaja izvirno v e-obliki in ga želimo v tej obliki tudi ohranjati ter kasneje predajati v pristojne arhive, kot to določa tudi ZVDAGA (40. člen, 4. odstavek). V naslednjem obdobju načrtujemo izvedbo delavnic, kjer bomo na primerih iz prakse natančno preizkusili najdena programska orodja za pretvorbo e-pošte po vseh relevantnih kriterijih. Ker smo že v sedanjih raziskavah ugotovili, da obstaja še vrsta drugih omejitev in neznank pri dolgoročnem ohranjanju e-pošte, bomo poiskali ustrezne odgovore tudi na njih. Eno takšnih omejitev najdemo v okoljih poštnih odjemalcev, ki omejuje velikost izvoznih map, drugi izziv predstavljajo priponke na drugih lokacijah itd. Na podlagi pridobljenih rezultatov bomo izvedli selekcijo tistih programskih orodij, za katere bomo ocenili, da so najprimernejša, tako z vidika pričakovanih funkcionalnosti kot širše dostopnosti. Za izbrana programska orodja bomo napisali navodila, ki bodo ustvarjalcem omogočila dolgoročno ohranjanje e-pošte in njeno ustrezno pripravo za kasnejši prevzem v sistem e-ARH.si, slednja bomo ponovno preverili s pilotskimi prevzemi e-pošte z izbranimi ustvarjalci.

SUMMARY

TECHNICAL ASPECTS OF LONG-TERM E-MAIL PRESERVATION

Klavdija Krivec

Archives of the Republic of Slovenia, Ljubljana, Slovenia
klavdija.krivec@gov.si

Tatjana HAJTNIK, Ph. D.

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

The e-ARH project team is preparing a solution for long-term e-mail preservation as a part of further development of the e-ARH.si project, based on the Strategy and Implementation Plan for the Development of the Slovenian Electronic Archives for the period from 2016 to 2020. Initial results of email preservation from the technical point of view are presented. The goal is to establish a stable and trusted Slovenian Electronic Archive e-ARH.si for email. The article summarizes email format issues, the possibility of transformation into archival formats and their validation. According to the carried out survey, the authors focused on an email solution for the most frequently used email applications of our creators, IBM Lotus Notes and Microsoft Outlook. Possible scenarios of exports and conversions were designed for further decisions in the selection of appropriate technical solutions for the long-term email preservation. This is followed by the implementation of workshops for testing the found software tools and choosing the most appropriate in accordance with the desired functionality and wider availability. After the selection of appropriate software, the team will create instructions for creators. Based on the strategy and expertise, acquired knowledge and best practices, we will enable a reliable and efficient long-term email preservation in the project plan for e-ARH.si 2016-2020.

Viri in literatura

- Hajtnik, T. (2011).** Strategija razvoja slovenskega javnega elektronskega arhiva "e-ARH.si". V: *Knjižnica*, 1, 39-56. Ljubljana: Zveza bibliotekarskih društev Slovenije. Pridobljeno s: <https://www.dlib.si/details/URN:NBN:SI:DOC-W49HEO29/>.
- Hajtnik, T. (2012).** Analiza primernosti formata PDF/A za dolgoročno hrambo gradiva v digitalni obliki. V: *Arhivi, letn.* 35 (1), str. 157–182. Ljubljana: Arhivsko društvo Slovenije.
- Industry Supported PDF/A Validation (2017).** Pridobljeno 1. 3. 2018 s spletne strani: <http://verapdf.org/home/#validation>.
- ISO 14721:2012: Space data and information transfer systems – Open archival information system – Reference model (2012).** Geneva: International Organization for Standardization.
- Prom C. J. (2011).** *Preserving Email*. DPC Technology Watch, 4–8. Pridobljeno 5. 8. 2017 s <https://www.dpconline.org/docs/technology-watch-reports/739-dpctw11-01-pdf/file-01>.

Strategija in izvedbeni načrt razvoja slovenskega elektronskega arhiva 2016 – 2020. (2016). Pridobljeno 5. 12. 2017 s
[http://www.arhiv.gov.si/fileadmin/arhiv.gov.si/
pageuploads/zakonodaja/Strategija_e-ARH.si/Strategija_e-ARH_si_2016-
2020_1.0.pdf](http://www.arhiv.gov.si/fileadmin/arhiv.gov.si/pageuploads/zakonodaja/Strategija_e-ARH.si/Strategija_e-ARH_si_2016-2020_1.0.pdf).

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.