



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: 17. 04. 2018

1.02 Pregledni znanstveni članek

1.02 Review Article

ALI NAM LAHKO PRI VREDNOTENJU IN ODBIRANJU ELEKTRONSKEGA GRADIVA POMAGA TEHNOLOGIJA?

dr. Tatjana HAJTNIK

Arhiv Republike Slovenije, Ljubljana, Slovenija

tatjana.hajtnik@gov.si

mag. Aida Škoro Babić

Arhiv Republike Slovenije, Ljubljana, Slovenija

aida.skoro@gmail.com

Izvleček:

V prispevku avtorici preučujeta vprašanje, ali nam pri vrednotenju in odbiranju elektronskega gradiva lahko pomaga tehnologija. Pri tem izhajata iz dejstva, da vedno več dokumentarnega gradiva izvirno nastaja v elektronski obliki, kar posledično vpliva na vprašanje postopkov vrednotenja in odbiranja arhivskega gradiva. Pri ohranjanju elektronskega gradiva v Sloveniji arhivisti opravljajo pomembno vlogo že pri samem nastajanju gradiva. Ker vrednotenje zapisov temelji na številnih merilih, vključno z njihovo zgodovinsko, pravno, upravno, kulturno in znanstveno vrednostjo, v prispevku avtorici preučujeta in opredeljujeta, v katerih fazah postopkov vrednotenja in odbiranja bi lahko uporabili programska orodja. Zagovarjata trditev, da bo uporaba tehnologije za strojno vrednotenje in odbiranje edina možnost, ki bo v prihodnosti zagotavljala učinkovito delo arhivistov pri obvladovanju obilice e-zapisov. Preko opredelitve osnovnih pojmov, pregleda obstoječe literature in dobrih praks ter analizi dosedanjih rešitev na področju strojnega vrednotenja in odbiranja bosta avtorici poskusili odgovoriti na vprašanje, ki je zastavljeno v naslovu prispevka in iz katerega sta izhajali na začetku raziskave. Prispevek bo podal možne okvirje in znotraj njih uporabne rešitve pri vprašanju strojnega vrednotenja in odbiranja.

Ključne besede:

e-zapisi, arhivsko gradivo, avtomatično vrednotenje, avtomatično odbiranje, uporaba programskih orodij, tehnično vrednotenje, strojno vrednotenje

Abstract:

Can Technology Help the Process of Appraisal and Selection of E-Records?

The authors examine the issue, whether technology can help in the records appraisal and selection process. Starting point is the fact that the number of records in electronic form is increasing every day, which consequently influences the question of the procedures of appraisal and selection of archival records. In preserving electronic material in Slovenia, archivists play an important role in the creation of the material itself. Because the records appraisal is based on a number of criteria, including the historical, legal, administrative, cultural and scientific value of records, the authors examine and determine the stages in the procedures of appraisal and selection of records, in which software tools could be used. The paper argues that the use of technology for machine appraisal and selection will be the only option in the future that will ensure the efficient work of archivists in managing a plethora of e-records. Through the definition of basic concepts, the review of existing literature and good practices, the analysis of the previous solutions in the field of machine appraisal and selection, the authors will try to answer the question

posed at the beginning of the research. The contribution will provide possible frameworks and within them a useful solution to the issue of machine evaluation and selection.

Key words:

e-records, archival records, automated appraisal, automated selection of archival records, use of program tools, technical appraisal, machine appraisal

1. Uvod

Zaradi velikih količin zapisov oz. dokumentarnega gradiva¹, ki jih javnopravne osebe² vsakodnevno ustvarijo pri poslovanju, predstavlja vrednotenje (angl. *appraisal*), na podlagi katerega se določa, kateri zapisi imajo arhivsko vrednost, že več let eno najpomembnejših nalog javnih arhivov po vsem svetu. Vrednotenje zapisov temelji na številnih merilih, vključno z njihovo zgodovinsko, pravno, upravno, kulturno in znanstveno vrednostjo. V Sloveniji je s predpisi določeno, da se po izvedenem vrednotenju arhivsko gradivo odbira iz dokumentarnega gradiva javnopravnih oseb v skladu s strokovnimi navodili pristojnega arhiva in dodatnimi strokovnimi navodili predstavnikov pristojnega arhiva ob samem odbiranju (angl. *selection*) (ZVDAGA, 40. člen).

Ker vedno več dokumentarnega gradiva nastaja izvirno v elektronski obliki, postajata njegovo vrednotenje in odbiranje arhivskega gradiva iz dokumentarnega hkrati tudi eden največjih izzivov javnih arhivov. Tega se je zavedal tudi zakonodajalec, ki je z novelo Zakona o varstvu dokumentarnega in arhivskega gradiva ter arhivih (ZVDAGA, 2014) določil, da morajo pristojni javni arhivi v Sloveniji svojim ustvarjalcem v primeru izročitve elektronskega gradiva izdati dodatna strokovno-tehnična navodila najkasneje ob začetku odbiranja arhivskega gradiva kot del pisnih strokovnih navodil, s katerimi predpišejo tudi obseg, postopek in način izročitve elektronskega gradiva ter oblike in nosilce zapisov. Pri ohranjanju elektronskega gradiva v Sloveniji arhivisti opravljajo pomembno vlogo že pri samem nastajanju gradiva, saj že omenjeni zakon določa, da mora ustvarjalec z vsem gradivom ravnati kot z arhivskim, dokler mu pristojni arhiv ne izda pisnih strokovnih navodil (ZVDAGA, 34. člen).

Slovenska javna arhivska služba (SJAS)³ do sedaj vrednotenju zapisov v elektronski obliki (v nadaljevanju *e-zapisi*) ni namenjala toliko pozornosti, kot jo sicer namenja vrednotenju dokumentarnega gradiva v fizični obliki⁴. Zanimivo je, da tudi standard ISO 14721⁵, znan kot referenčni model odprtega arhivskega informacijskega sistema (OAIS), ki danes predstavlja ključni standard na področju ohranjanja e-zapisov

¹ Dokumentarno gradivo so vse vrste in oblike zapisov, ki so nastali ali bili prejeti pri poslovanju pravnih in fizičnih oseb (ZVDAGA, 2. člen).

² Javnopravne osebe so državni organi, samoupravne lokalne skupnosti ter pravne osebe javnega prava in zasebnega prava ter fizične osebe, ki so nosilci javnih pooblastil ali izvajalci javnih služb.

³ Slovensko javno arhivsko službo predstavljajo Arhiv Republike Slovenije in šest regionalnih arhivov (Zgodovinski arhiv Ljubljana, Pokrajinski arhiv Maribor, Zgodovinski arhiv na Ptuju, Zgodovinski arhiv Celje, Pokrajinski arhiv Nova Gorica, Pokrajinski arhiv Koper).

⁴ Dokumentarno gradivo v fizični obliki je dokumentarno gradivo na fizičnem nosilcu zapisa (npr. na papirju, filmu itd.), ki omogoča reprodukcijo vsebine brez uporabe informacijsko komunikacijskih ali sorodnih tehnologij (ZVDAGA, 2. člen).

⁵ Standard ISO 14721 je bil leta 2013 sprejet kot slovenski standard, SJAS ga je skupaj z Narodno univerzitetno knjižnico in pod pokroviteljstvom Ministrstva za kulturo prevedla v slovenski jezik (SIST ISO 14721, 2013).

po vsem svetu in tudi v Sloveniji, vrednotenja zapisov v svojem funkcionalnem modelu eksplicitno ne vključuje kot funkcijo ali podfunkcijo, temveč ga omenja le kratko:

»Organizacija, ki upravlja OAIS, bi morala vpeljati nekatera merila za določanje vrste informacij, ki jih je pripravljena sprejeti (oz. se od nje zahteva, da jih sprejme). Ta merila lahko vključujejo temo informacij, vir informacij, stopnjo unikatnosti ali originalnosti, obliko tehnologije, uporabljene za predstavitev informacij (t. j. fizični medij, digitalni medij, format).« (ISO, 2013, poglavje 3.2.1)

Enako tudi standard ISO 20652 (2006), ki podrobno opisuje sodelovanje med ustvarjalci arhivskega gradiva in arhivi, izrecno ne vključuje postopka vrednotenja zapisov pred njihovo predajo v arhivsko hrambo. Ne glede na to so se arhivi po svetu (in upravljalci gradiva) začeli ukvarjati s problemom vrednotenja e-zapisov že pred nekaj desetletji ter v ta namen ustvarili vrsto teorij in metod.⁶ V okviru projekta InterPARS1 (2001) so ugotavljali, ali se teorija in metodologija vrednotenja e-zapisov razlikujeta od tistih za gradivo na fizičnem nosilcu zapisa in kakšno vlogo ima postopek vrednotenja na dolgoročno ohranjanje e-zapisov. Proučevali so, ali imajo tehnologija, nosilci in oblike zapisa kakšen vpliv na postopke vrednotenja, kdaj bi bilo treba vrednotiti e-zapise, ali bi bilo morda treba postopke vrednotenja e-zapisov v času njihovega obstoja večkrat ponoviti, in če da, kdaj in kdo naj bi bil odgovoren za vrednotenje e-zapisov, ter ne nazadnje, kateri so kriteriji in metode vrednotenja verodostojnega e-zapisov.

SJAS se je z vsemi temi vprašanji začela ukvarjati pred nekaj leti, in sicer v času priprave prve strategije razvoja slovenskega elektronskega arhiva (v nadaljevanju e-ARH.si), ko je v Izvedbeni načrt omenjene strategije umestila tudi projektno nalogo z naslovom *»Avtomatizirano (strojno) vrednotenje prevzetega digitalnega arhivskega gradiva (PN II/3.1.6)«* (Arhiv RS, 2010a). Ključni namen te naloge je bil preveriti tiste podatkovne strukture in tudi vsebine, ki niso relevantne ali pomembne v okviru celot, ki jih SJAS prevzema kot arhivsko gradivo, s ciljem izdelave programskega orodja za (strojno) vrednotenje posamezne vrste prevzetega elektronskega arhivskega gradiva. Strategijo (Arhiv RS, 2010b) in izvedbeni načrt je leta 2010 potrdila tudi Vlada RS (Sklep vlade, 2010), vendar pa za samo operativno izvedbo ni zagotovila zadostnih kadrovskih in finančnih virov, zato naloga ni bila realizirana. Je pa ta ponovno uvrščena v strategijo in izvedbeni načrt za razvoj e-ARH.si za obdobje 2016–2020 (Arhiv RS, 2016), ki jo je prav tako potrdila Vlada RS (Sklep vlade, 2016). Za uresničitev strategije je bil v Arhivu RS vzpostavljen projekt, ki je uspešno pridobil finančna sredstva iz Operativnega programa za izvajanje Evropske kohezijske politike v obdobju 2014–2020. Namen in cilj naloge z naslovom *»Razvoj oz. nabava programske opreme za avtomatizirano vrednotenje elektronskega arhivskega gradiva«* ostajata ista, za njuno dosego pa je kot prva aktivnost predviden pregled dobrih praks na tem področju.

Ta prispevek predstavlja praktično prvi korak k realizaciji zgoraj omenjene strateške naloge, pridobljene ugotovitve pa bodo služile za nadaljnje delo. Najprej opredelimo e-zapis in njegove bistvene značilnosti ter izpostavimo ključne izzive pri njihovi dolgoročni hrambi, nato kratko predstavimo postopka valorizacije in odbiranja zapisov z arhivsko vrednostjo, v zaključnem delu pa predstavimo dobre prakse na področju vrednotenja in odbiranja e-arhivskega gradiva iz množice e-zapisov ter prve odgovore na to, ali lahko arhivistom pri teh postopkih pomaga tehnologija oz. ali jih je mogoče vsaj delno izvesti s pomočjo tehnologije.

⁶ Npr. Schellenbergerjev (1956) model vrednotenja vladnega gradiva ali projekt InterPARES1 (2001).

2. E-zapis

Raziskovalci arhivisti univerze University of British Columbia so v sodelovanju z ameriškim Department of Defense v okviru raziskovalnega projekta *The Preservation of the Integrity of Electronic records*, ki so ga izvajali v obdobju med 1994 in 1997, definirali e-zapis (angl. *electronic record*) kot zapis, ustvarjen v elektronski obliki, ki ga v informacijskih sistemih najdemo v obliki dokumenta, informacije ali podatka (Duranti, 2001). Pri tem so podatke opredelili kot najmanjše nedeljivo dejstvo, informacije kot dano znanje ali sporočilo, namenjeno komunikaciji, in dokument kot zapisano informacijo.

Zapis na fizičnih nosilcih je fiksni in ostane nespremenjen ves čas njegove hrambe. Glede na dosedanje izkušnje to lahko traja zelo dolgo časa, saj npr. papir velja za razmeroma trajen nosilec in je treba relativno malo za njegovo ohranjanje. Zadostuje že hramba v prostoru z ustrežno temperaturo, vlago in svetlobo, zaščita pred ognjem in nepooblaščenimi dostopi osebam, ki bi lahko povzročile njegovo poškodovanje ali uničenje. Tudi uporaba zapisov na fizičnih nosilcih je skozi daljše časovno obdobje dokaj enostavna. Preprosto jih vzamemo v roke in njihova vsebina je človeškemu očesu takoj dostopna.

Pri e-zapisih pa takšen način hrambe ne zadostuje, saj jih ni mogoče prebrati, brez da bi imeli na voljo ustrezno tehnologijo (strojno in programsko opremo), ki bo prevedla strojni zapis (niz bitov) v človeku razumljivo obliko. Zato je povezanost dostopa do e-zapisov in njihovega dolgoročnega ohranjanja dosti večja kot pri zapisih na fizičnih nosilcih, saj se dostop do e-zapisov zaradi hitrega tehnološkega razvoja lahko izgubi v relativno kratkem času, če ne sprejmemo ustreznih aktivnosti za njihovo ohranitev že vse od njihovega nastanka.

E-zapisi se lahko, za razliko od tistih na fizičnih nosilcih, hitreje in neopaženo spremenijo. Tako že vrsto let enega večjih izzivov predstavlja ohranjanje e-zapisov na način, da ohranimo njihovo enakost *izvirnikom* in s tem tudi njihovo *verodostojnost*⁷.

3. Izvirnik e-zapisa

Za razliko od zapisa na fizičnem nosilcu pri e-zapisu ni povsem razjasnjeno, kaj predstavlja njegov izvirnik. Pri zapisu na fizičnem nosilcu (npr. papirju) je običajno to enostavno določiti. Ob nastanku je praviloma opremljen z določenimi podatki, ki ga opredeljujejo kot izvirnik, npr. lastnoročni podpis, datum, naslovnik, pošiljatelj, morda logotip inštitucije. Vsaka kopija zapisa na papirju je reprodukcija izvornika in to običajno zaznamo že s prostim očesom. S pravnega vidika je za priznavanje istovetnosti kopije zapisa na papirju z njegovim izvirnikom potrebna overitev oz. potrditev s strani pooblaščenih oseb (npr. overitev na upravni enoti v upravnih postopkih ali pri notarju). Razliko med zapisom na fizičnem nosilcu, npr. dokumentom na papirju, in e-dokumentom zelo enostavno obrazloži Lynch (2000). Pravi, da izvirnik dokumenta na papirju predaš in ga nimaš več pri sebi, medtem ko e-dokument predaš, hkrati pa ga imaš lahko še vedno pri sebi. Pri e-dokumentu je torej zelo enostavno narediti več njegovih kopij, težje pa je določiti, kateri od teh dokumentov je izvirnik. Levy (2000) na primer dokazuje, da so vsi e-dokumenti dejansko kopije, vprašanje je samo, ali imajo vse kopije nekega dokumenta enake lastnosti izvornika. Rothenberg (2000) pa zagovarja nasprotno stališče in pravi, da je vsaka upodobitev dokumenta »*elektronski izvirnik*« takrat, ko ohranja kar največ smiselnih in bistvenih lastnosti izvornika. Zagovarja tudi, da

⁷ Verodostojnost predstavlja dokazljivost povezanosti reproducirane vsebine z vsebino izvornega gradiva oziroma izvorom tega gradiva ali poenostavljeno povedano, gradivo ne sme biti nepooblaščenno spremenjeno, ves čas hrambe mora biti enako izvorniku.

e-dokumentu lahko zagotavljamo vrednost izvirnika z zanesljivimi kontrolami celovitosti in zanesljivosti. Kot primer navaja prevzem e-dokumenta v arhivsko ustanovo, kjer zanjo prevzeti e-dokument od ustvarjalca enostavno predstavlja prvi izvirnik.

4. Bistvene lastnosti e-zapisov

Ne glede na nosilec (fizični ali elektronski) ima vrednost zapisov odločilno vlogo pri določitvi, kako dolgo jih bomo hranili. Glede na izzive, povezane predvsem s hitrim tehnološkim razvojem, možnostjo enostavnega spreminjanja e-zapisov ter na drugi strani zahtevnosti ugotavljanja in dokazovanja teh sprememb, se postavlja ključno vprašanje, ali bodo e-zapisi čez čas še dostopni in bodo ohranjali vse *bistvene lastnosti izvirkov*, s tem pa tudi njihovo vrednost. Uredba o varstvu dokumentarnega in arhivskega gradiva (2017, 2. člen) definira, da je izvirnik vsak primaren oziroma izviren zapis ne glede na pisno osnovo in način zapisa. Če izhajamo iz tega, da so bistvene lastnosti zapisov tiste, ki so ključne za njihov pomen, uporabo ali organizacijsko vrednost, potem jih je treba ohraniti tudi pri kakršni koli posegih v njih, prenašanju v druga tehnološka okolja ali pretvarjanju v druge formate. Nekaj osnovnih virov kot bistvene lastnosti e-zapisov izpostavljajo njihovo *vsebino* (angl. *content*), *kontekst* (angl. *context*), *strukturo* (angl. *structure*) ter *videz in obnašanje* (angl. *look and feel*) (Hajtnik, 2016).

V primerjavi z zapisi na fizičnih nosilcih lahko pride pri e-zapisih do izgube njihovih bistvenih lastnosti že, če jih iz informacijskega okolja, v katerem so bili ustvarjeni, prenesemo in predstavljamo oz. uporabljamo v drugih informacijskih okoljih. Posledično dostop do teh e-zapisov v celoti ni več mogoč oziroma, v kolikor do njih dostopamo s sodobnejšimi programskimi orodji, lahko pride do nepričakovane izgube dela njihovih bistvenih lastnosti. Posledično sledi vprašanje, v kolikšni meri je še zagotovljena istovetnost teh zapisov z izvirkami in s tem zaupanje v njihovo verodostojnost.

Bearman (1999) pravi, da sta zlasti *vsebina* in *kontekst* tisto, s čimer je zapis opredeljen kot dokaz, tj. umeščenost v poslovno aktivnost in njegovo upravljanje. *Kontekst* se nanaša na izvorno okolje, v katerem je e-zapis nastal in se je uporabljal. Za kontekstualno razumevanje e-zapisa potrebujemo določeno količino podatkov o njegovem nastanku, na primer kdo in kdaj ga je ustvaril in v katerem poslovnem procesu. Kontekst je predstavljen kot dodatna informacija, ki ponazarja odnos med zapisom in drugimi zapisi ter aktivnost, v okviru katere je bil zapis oblikovan (ICA, 1997). Zato je kontekst e-zapisa lahko predstavljen z notranjimi (npr. povezave, priloge, reference) ali zunanji elementi (npr. metapodatki), ki omogočajo njegovo razumevanje, ga povezujejo z administrativnim in funkcionalnim okoljem ter drugimi zapisi. *Struktura* e-zapisa je vezana na način, kako je zapisan, predstavljata jo logična hierarhija in odnos med njegovimi različnimi deli. Elementi strukture zapisa (npr. dokumenta) so lahko npr. naslovnica, poglavja, odstavki, bibliografija ali priloge, opombe in komentarji. Izguba elementov, ki opredeljujejo strukturo e-zapisa, lahko povzroči njegov nepravilen prikaz ali napačno interpretacijo njegove vsebine. *Vsebina* e-zapisa predstavlja golo besedilo (tekst), ki je neodvisno od strukture (npr. poglavij ali odstavkov) ter njegovega videza in obnašanja (npr. vrste ali velikosti pisave, barve pisave, položaja številčenja strani). *Videz in obnašanje* e-zapisa pa se nanaša na njegovo končno predstavitev, torej na obliko, v kateri ga vidimo na računalniškem zaslonu.

Bistvene lastnosti e-zapisov opisujemo z metapodatki, ki jih k e-zapisom dodajamo kot dodatne informacije. So ključni za zagotavljanje, da bodo e-zapisi »preživel« in bodo na voljo tudi v prihodnosti. Za potrebe opisa bistvenih lastnosti e-zapisov obstaja več različnih metapodatkovnih shem, strategij in standardov (npr. METS⁸, Dublin Core⁹, PREMIS¹⁰). Katere od lastnosti e-zapisov, opisane z metapodatki, so pomembnejše oz. sodijo med bistvene, je odvisno tudi od samega tipa e-zapisov (npr. e-pošte, tekstovnih dokumentov, preglednic in podatkovnih baz), v prvi vrsti so določene s posameznim poslovnim procesom (Hajtnik, 2016). Tako na splošno lahko predvidevamo, da bo videz in obnašanje e-poštnih sporočil manj pomemben, saj je znano, da se predstavnost e-pošte razlikuje med osebnimi računalniki, ki uporabljajo različne programe za e-pošto in imajo različne osebne nastavitve. Nasprotno pa je lahko za e-dokumente videz in obnašanje bistvenega pomena. Lastnosti e-zapisov, ki nimajo bistvenega pomena za njihovo sporočilnost, tudi nimajo bistvenega pomena za sam zapis. Sem na primer lahko štejemo značilnosti aplikacije, s katero so bili e-zapisi ustvarjeni. Namreč, če bi želeli ohraniti vse lastnosti e-zapisov, bi posledično lahko to pomenilo porabo velike količine prostora, v bistvu pa so nekatere, npr. z vidika zgodovinske ali arhivske vrednosti, nepomembne.

5. Vrednotenje in odbiranje

5.1 Pomen vrednotenja in odbiranja

Pri obravnavanju vprašanja vrednotenja in posledično odbiranja arhivskega gradiva je prav, da razumemo pomen obeh dejanj oziroma procesov, posebno zato, ker gre za dva tesno povezana, vendar različna procesa. **Odbiranje** je na praktičen način udejanjanje rezultatov oziroma odločitev zaključenega vrednotenja. **Vrednotenje** je namreč dejanje intelektualne odločitve, ki je predpogoj oziroma predhodnica odbiranja (Harrison, 1997). Vrednotenje je tako intelektualni okvir, znotraj katerega se izvede odbiranje. Torej govorimo o merilih za ohranitev arhivskega gradiva, in sicer z vidika pomena gradiva za posameznike, skupino posameznikov, lokalno in državno skupnost ter tudi širše. O pomenu gradiva v tem kontekstu odloča vsaka država posebej (Škoro Babić, 2017).

5.2 Zakaj vrednotiti?

Potreba po vrednotenju gradiva se je razvila istočasno z nastankom arhivov. Količine zapisov so bile namreč v razmerju do prostorov, predvidenih za shranjevanje, mnogo večje. Vseh zapisov res ni treba ohraniti, če lahko izmed njih pravilno in pravično odberemo tiste, ki bodo služili za poznejše potrebe. V ta namen pa je treba opredeliti vsebinska merila, s pomočjo katerih bi se lahko pravilno odločilo, kateri zapisi imajo trajno vrednost. Vpogled v zgodovino vrednotenja zapisov na podlagi ohranjenega gradiva nam kaže, da je bilo arhivsko gradivo ohranjeno predvsem zaradi pravne

⁸ METS - Metadata Encoding, Transmission Standard: je shema XML za definiranje opisnih, administrativnih in strukturnih metapodatkov. Uporabna je za izmenjavo podatkov med sistemi za hrambo e-zapisov, za shranjevanje podatkov o e-zapisih v posameznih zbirkah, npr. v knjižnicah, arhivih in muzejih, ali za prikaz rezultatov iskanja po teh zbirkah na spletu skupaj s samimi e-zapisi.

⁹ Dublin Core je standard metapodatkovnih elementov s 15 lastnostmi za opisovanje virov, preko katerih se te opišejo in katalogizirajo za lažje iskanje virov. Dublin Core lahko opisuje fizične lastnosti knjige, e-zapise in tudi spletne strani.

¹⁰ PREMIS Data Dictionary for Preservation Metadata je v svetu široko uporabljen podatkovni slovar za ohranjanje metapodatkov, saj je njegova podpora vključena v številna komercialna in odprtokodna orodja ter sisteme za e-hrambo. Opredeljuje nabor metapodatkov za podporo dolgoročne hrambe e-zapisov glede na njihovo vsebino in strukturo, izvajalce ter procese dolgoročne hrambe. Je mednarodni standard in izhaja iz standarda ISO 14721.

varnosti političnih tvorb oziroma nosilcev oblasti. S francosko revolucijo se je pomen arhivskega gradiva in arhivov preoblikoval, arhivi so postopoma postali varuhi zgodovinskega in kulturnega spomina, identitete narodov in držav ter človekovih pravic.

Prelomnico v zgodovini vrednotenja zagotovo predstavlja delo Hilary Jenkinson »*A Manual of Archive Administration*« (1922), v katerem je utemeljeval, da je glavna dolžnost arhivistov, da zapis ohranijo, nato pa sledi dolžnost do uporabnika (Jenkinson, 1922). Ta misel je tudi po skoraj stotih letih še kako aktualna, ko govorimo o gradivu v elektronskem okolju, kjer v prvi vrsti ohranitev avtentičnosti¹¹ in celovitosti¹² zapisa ob ohranitvi konteksta predstavlja velik izziv.

Slovenska arhivska stroka se je naslonila na Schellenbergovo teorijo moderne valorizacije, ki je vsebovala začetke funkcijskega vrednotenja ter temeljna načela in kriterije vrednotenja (Schellenberg, 1956/1984). Ob upoštevanju nemške teorije, kjer je treba omeniti Karla Otta Müllerja in Heinricha Otta Messnerja, ki sta utemeljevala merila za vrednotenja skozi kontekst struktur inštitucij (Kolsrud, 1992), v okviru katerih gradivo nastaja, ter ob upoštevanju sovjetske teorije, ki je Schellenbergov vsebinski model še dodelala (Škoro Babić, 2017), so v slovenskem prostoru začeli nastajati klasifikacijski načrti pri ustvarjalcih, ki so predstavljali in še danes predstavljajo osnovni pripomoček za makrovalorizacijo. Klasifikacijski načrt je sistem razvrščanja dokumentov, zadev in dosjejev na podlagi vsebine. Predstavlja osnovni šifrant za razporejanje dokumentarnega gradiva in predstavlja konkretno podlago oz. pripomoček za vrednotenje in odbiranje arhivskega gradiva. Metodologija in oblikovanje meril za vrednotenje arhivskega gradiva je kompleksno vsebinsko vprašanje, ki ga na tem mestu ne bomo obravnavali, vendar je treba poudariti, da je za kakovostno vrednotenje potreben znanstvenoraziskovalni pristop, ki ga je nemogoče oziroma se ga ne sme nadomestiti z avtomatizmom. Vsekakor pa se določene faze določanja e-gradiva, ki ima značilnosti arhivskega, na podlagi znanstvenoraziskovalnega elaborata lahko avtomatizirajo oziroma izvajajo s pomočjo programskih orodij.

6. Pregled dobrih praks na področju vrednotenja e-zapisov

Vprašanje strojnega vrednotenja in odbiranja buri mednarodne arhivske kroge že več desetletij. Naugler (1984) je že davnega leta 1984 v študiji obdelal vprašanje vrednotenja e-zapisov, opredelil osnovne pojme ter izpostavil faktorje tveganja. Arhivska stroka se je pri vprašanju vrednotenja arhivskega gradiva osredotočala na vprašanje, kdaj izvesti vrednotenje. Cook (1992) je zagovarjal stališče, da vrednotenje e-zapisov ne sme biti drugačno od vrednotenja gradiva na klasičnih nosilcih. Z enakimi izhodišči so začeli snovalci projekta InterPares1 (2001).

Obstoječa literatura je seveda koristna, vendar pa je zaenkrat mogoče najti le malo nasvetov o praktični uporabi takšnih okvirov. Enega takšnih predstavlja Tajvanski nacionalni arhiv, ki izvaja tehnično vrednotenje e-zapisov že od leta 2008 (National Archives Administration, 2018). Njihovo izhodišče je, da je namen tehničnega vrednotenja e-zapisov preveriti verodostojnost in celovitost ohranjenih elektronskih zapisov, vključno z uspešnostjo njihovega fizičnega prenosa. Razvili so devet kazalnikov in pripadajočih validacijskih metod za tehnično vrednotenje e-zapisov: *avtentičnost in celovitost e-zapisov, učinkovitost nosilcev zapisa, vrste e-zapisov, vrste formatov*

¹¹ ZVDAGA (2014, 27. člen) tolmači avtentičnost kot dokazljivo povezanost reproducirane vsebine z vsebino izvirnega gradiva oziroma izvorom tega gradiva.

¹² ZVDAGA (2014, 5., 26. in 27. člen) tolmači celovitost kot nespremenljivost in integralnost dokumentarnega gradiva oziroma reprodukcije njegove vsebine, urejenost dokumentarnega gradiva oziroma njegove vsebine ter dokazljivost izvora dokumentarnega gradiva (provenience).

datotek, robustnost e-ovojnice, robustnost certificiranega aplikacijskega sistema, robustnost enkripcije, zgoščevalne vrednosti (hash) in e-podpisa.

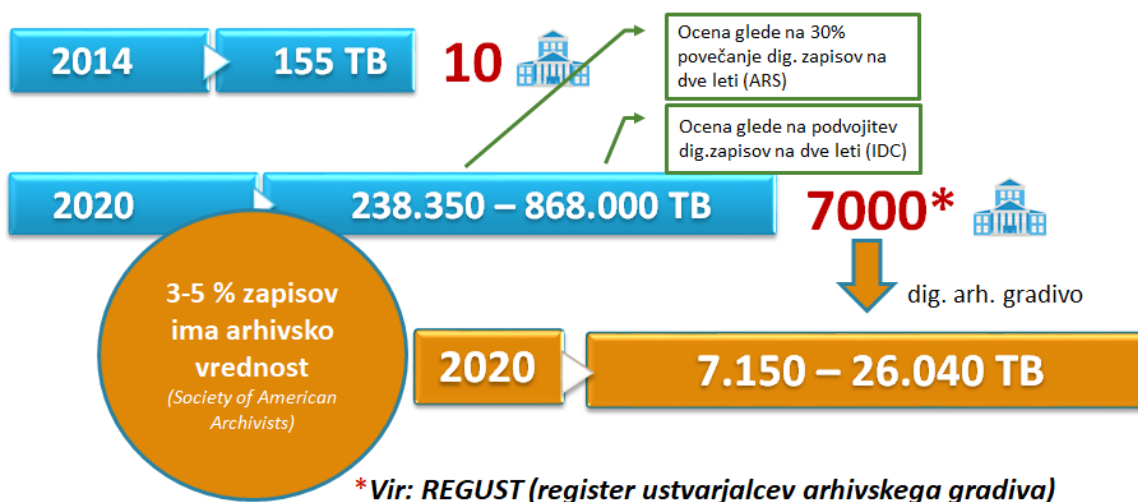
Velik korak pri avtomatizaciji procesov upravljanja z elektronskim gradivom ter posledično pri vrednotenju in odbiranju arhivskega gradiva so naredili v Združenih državah Amerike v Nacionalnem arhivu (NARA)¹³. Oblikovali so načrt upravljanja z dokumenti v elektronski obliki, s pomočjo katerega bi izboljšali interoperabilnost in s tem olajšali poslovanje vladnih agencij. Glede na podatke v letnem poročilu je razvidno, da ameriški nacionalni arhiv oblikuje celostno rešitev, ki bi se aplicirala v vseh zveznih organih in vladnih agencijah. Pristopili so k reševanju avtomatizacije procesov tako, da so ustanovili skupino strokovnjakov, ki bo oblikovala smernice in tehnične zahteve za razvijalce programske opreme, s katerimi bodo pridobili orodja, ki jim bodo ustrezala za avtomatizacijo procesov. Nekaj korakov pred ameriški arhivisti so zagotovo avstralski arhivski strokovnjaki, ki že nekaj let oblikujejo platformo elektronskega gradiva državne uprave in ki v arhiv ne sprejemajo več gradiva na klasičnih nosilcih.¹⁴ Ker tudi v avstralskem nacionalnem arhivu pristopajo k reševanju vprašanja elektronskega arhiviranja na celostni način, in sicer z vpeljavami tehničnih zahtev pri ustvarjalcih, trenutno rešitev, ki bi ustrezala vprašanju strojnega vrednotenja in odbiranja z vidika slovenske stroke, še ni. Glavni razlog za to je, da se koncept arhiviranja, ki ga imamo v Sloveniji in ki ga ima večina evropskih držav, zelo razlikuje od ameriškega in avstralskega koncepta, ki pa se tudi med seboj seveda razlikujeta. Obe državi uresničujeta vse arhivske zahteve že pri ustvarjalcu, torej v fazi, ko gradivo nastaja. Zaradi precej zahtevnega izziva reševanja vprašanj konkretne rešitve še niso dostopne. Velik kazalnik teh dosežkov je vsekakor mednarodni projekt InterPares1, ki združuje arhivske strokovnjake s celega sveta, vendar še ni moč zaslediti konkretne in celostne rešitve na področju avtomatizacije procesa valorizacije in odbiranja.

7. Avtomatiziranje postopkov vrednotenja in odbiranja e-zapisov

Mednarodna analitska hiša International Data Corporation (IDC, 2014) je na podlagi izvedene raziskave leta 2014 napovedala, da se bo količina e-zapisov od leta 2013 do 2020 vsaki dve leti več kot podvojila. Po raziskavah, ki jih navaja skupnost ameriških arhivistov (*Society of American Archivists, 2018*), naj bi imelo vsaj 3–5 % teh zapisov arhivsko vrednost. V Sloveniji je zaokroženo 7.000 ustvarjalcev arhivskega gradiva, izmed njih je 10 reprezentativnih ustvarjalcev leta 2014 skupaj ustvarilo 155 TB e-zapisov. Iz vseh teh podatkov lahko hitro ocenimo, da bodo leta 2020 v Sloveniji vsi ustvarjalci skupaj ustvarili 868.000 TB e-zapisov oz. 26.040 TB e-arhivskega gradiva (slika 1).

¹³ Načrt avtomatizacije procesov je dostopen na: <https://www.archives.gov/records-mgmt/prmd/automated-erm.html>.

¹⁴ Nacionalni arhiv Avstralije na svoji spletni strani pojasnjuje ključne točke digitalne strategije: <http://www.naa.gov.au/information-management/digital-transition-and-digital-continuity/information-is-managed-digitally/Whole-of-Government-Digital-Records-Platform.aspx> (25. 3. 2018).



Slika 1: Ocenjeno število digitalnih zapisov pri ustvarjalcih leta 2014–2020 v Sloveniji (Hajtnik, 2017a)

Za ilustracijo – če bi te zapise shranili na CD-je zmogljivosti 700 MB, bi jih za shranitev e-arhivskih zapisov leta 2020 potrebovali 39.007.027, če pa bi te zapise shranili na 4,7 GB zmogljive DVD-je, bi jih potrebovali 709.175 (slika 2).



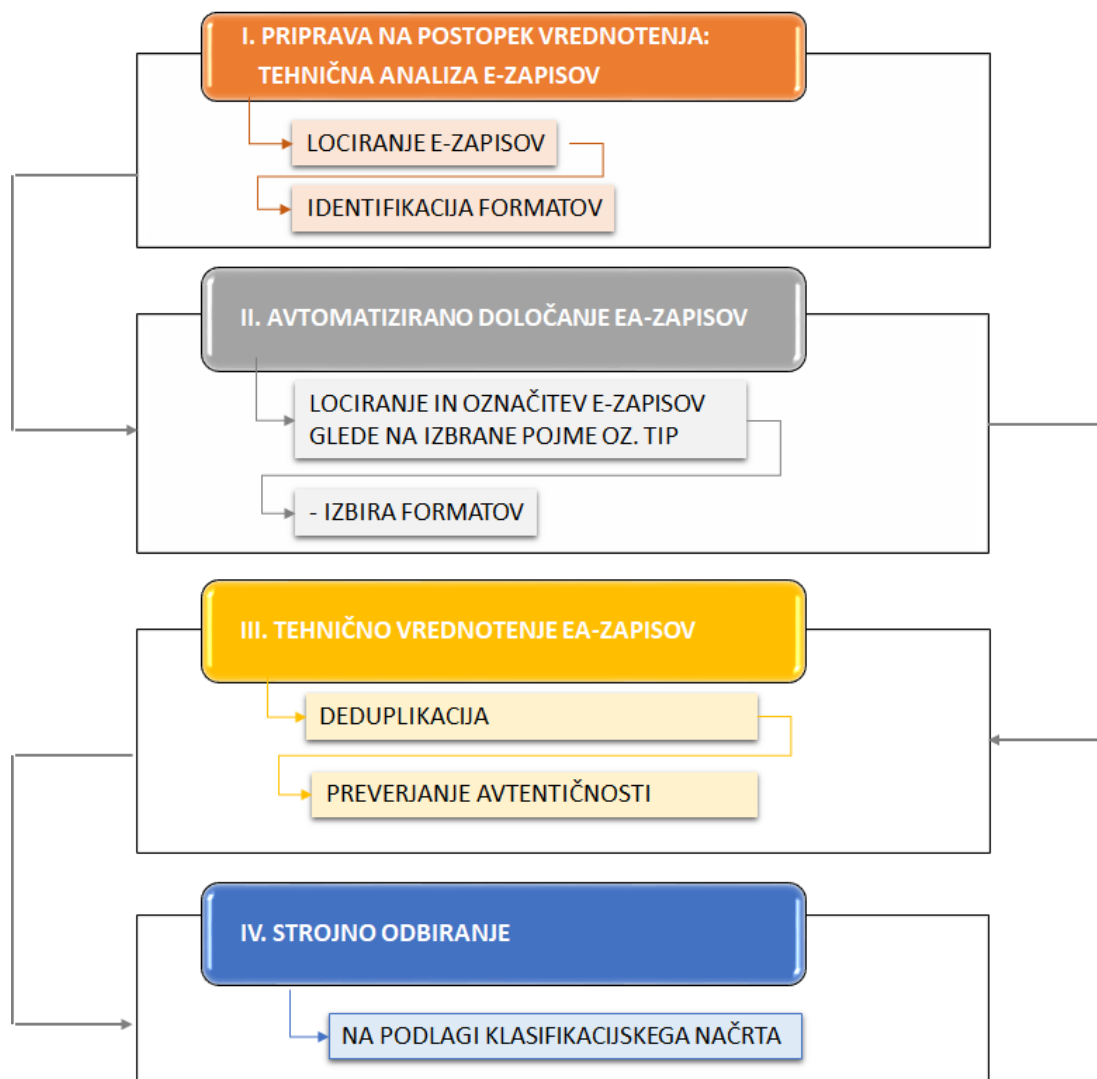
Slika 2: Ocenjeno število digitalnih zapisov izraženo s številom prenosnih medijev (Hajtnik, 2017a)

Omenjene količine e-zapisov bodo arhivistom prav gotovo predstavljale enega ključnih izzivov pri postopkih vrednotenja in odbiranja arhivskega gradiva (Hajtnik, 2017b). Zato bodo morali ponovno preučiti postopke vrednotenja in odbiranja arhivskega gradiva izmed množice e-zapisov pri svojih ustvarjalcih ter skupaj z njimi intenzivno sodelovati pri kreiranju politik, ki so namenjene izbrisu nepotrebnih e-zapisov, hkrati pa razmisliti, ali jim pri tem lahko pomaga tehnologija.

Avtomatiziranje vrednotenja gradiva je vsekakor in bo v prihodnosti vedno bolj aktualno vprašanje že zaradi množice e-zapisov. V Sloveniji in tudi drugod po svetu se je za zapise na klasičnih nosilcih (papirju) uveljavila praksa, da je treba izmed njih določiti tiste z arhivsko vrednostjo (arhivsko gradivo) čim bolj zgodaj v njihovem življenjskem krogu – že ob nastanku, seveda, v kolikor je to mogoče. Ker ima ohranjeno gradivo pričevalno vrednost le, če je ohranjeno v kontekstu (Škoro Babić, 2017), je uveljavljeno pravilo, da se k vrednotenju pristopa na znanstvenoraziskovalni način in s tem ustvarja intelektualni okvir. Ta pristop zajema preučevanje časa nastanka gradiva oziroma poslovanja ustvarjalca, pomen ustvarjalca v družbi v določenem časovnem obdobju, pomen gradiva, ki nastaja pri ustvarjalcu v določenem časovnem obdobju, in ne nazadnje preučevanje raziskovalnih trendov, s pomočjo katerih se valorizacijske odločitve dopolnjujejo (Škoro Babić, 2015). Slovenska zakonodaja tudi odlično rešuje vprašanje naknadne opredelitve nekega gradiva kot arhivskega, saj je resnično težko popolnoma natančno predvideti, kateri zapisi bodo imeli pomen za večno hrambo. Za zapise na fizičnih nosilcih velja, da prvotna faza v procesu vrednotenja ne more in ne sme biti prepuščena avtomatizmu.

Če ponovimo dejstvo, da vsakodnevno nastaja velika množica e-zapisov različnih tipov (besedilni dokumenti, audio/video zapisi, film, podatkovne baze, zapisi v informacijskih sistemih za upravljanje z dokumenti, e-pošta, spletne strani, družbena omrežja itd.), je (in bo vedno bolj) postopek vrednotenja in odbiranja nemogoče opraviti celovito in natančno ter v sprejemljivih časovnih rokih. Zato je skrajni čas, da začnemo razmišljati o tem, kako bi lahko pri tem čim bolj koristno uporabili tehnologijo oziroma namenska programska orodja. Ko začnemo razmišljati o vključevanju tehnologije v postopke vrednotenja in odbiranja, postane tudi jasno, da ti postopki ne bodo mogli biti več samo v domeni in izvajanju arhivistov, nujno bo potrebno sodelovanje informatikov, po vsej verjetnosti pa v določenih primerih tudi pravnikov.

Koncept vrednotenja arhivskih e-zapisov ne bi smel biti bistveno drugačen od tistega, ki velja za zapise v fizični (praviloma papirni) obliki, saj ga načeloma podpirajo iste arhivske teorije. Vendar pa je pri e-zapisih vendarle treba upoštevati, da so ti odvisni od tehnologije skozi celoten življenjski cikel. Kot bomo prikazali v nadaljevanju, na eni strani to dejstvo bistveno vpliva na sam pristop k postopku vrednotenja in odbiranja arhivskih e-zapisov, po drugi strani pa lahko ravno tehnologijo koristno uporabimo pri izvajanju teh postopkov. Na sliki 3 je prikazan predlog osnovnega koncepta pristopa k postopku vrednotenja in odbiranja arhivskih e-zapisov, ki ga bomo v nadaljevanju podrobneje razdelali.



Slika 3: Ključne faze pri avtomatiziranju postopkov vrednotenja in odbiranja e-zapisov

7.1 Priprava na postopek vrednotenja: tehnična analiza e-zapisov

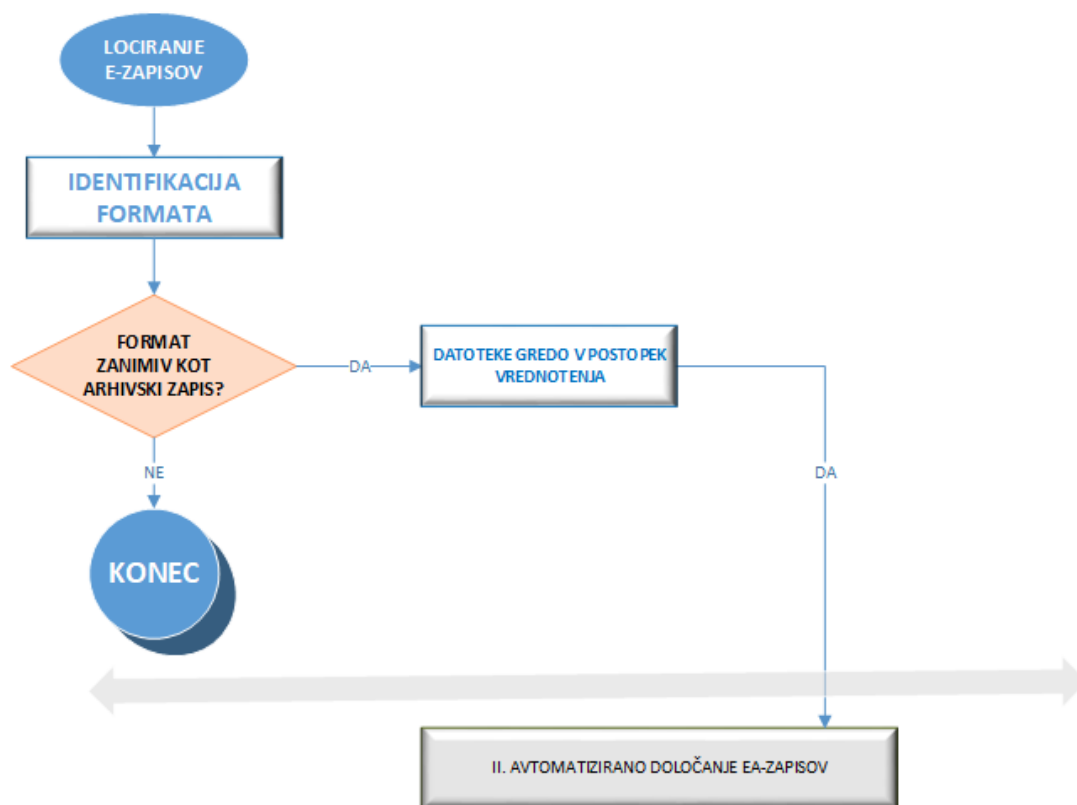
Pred začetkom postopka vrednotenja e-zapisov je smiselno, da najprej izvedemo tehnično analizo e-zapisov, to sestavljata dve ključni aktivnosti: *lociranje e-zapisov in identifikacija formatov*.

7.1.1 Lociranje e-zapisov

Prvi korak pred začetkom vrednotenja in odbiranja predstavlja *lociranje e-zapisov* (npr. zapisi v ISUD, podatkovni bazi, na datotečnem sistemu, prenosnih medijih) in ugotovitev, ali so dostopni. Prav hitro bi se namreč lahko zgodilo, da so e-zapisi shranjeni na nosilcih, ki s tehnologijo, ki je danes v širši uporabi, niso več nedostopni (npr. ker so zapisani na disketah ali CD-jih, pa nimamo ustreznih bralnih enot).

7.1.2 Identifikacija formatov

V računalniških sistemih obstaja množica e-zapisov v različnih oblikah zapisov (formatih). Zato je smiselno, da potem, ko smo locirali e-zapise, ki jih želimo vrednotiti, najprej ugotovimo, v katerih formatih so ti zapisani. Že iz samega formata namreč lahko ugotovimo, da e-zapis ni zanimiv za nadaljnje vrednotenje, ker predstavlja npr. sistemsko datoteko ali pa datoteko, ki predstavlja del programskega orodja. Te zapise lahko iz nadaljnega obravnavanja odstranimo (slika 4).



Slika 4: Ugotavljanje ali je e-zapis zanimiv za postopek vrednotenja glede na njegov format

Nekatera od programskih orodij identificirajo format na podlagi priponke (npr. File-Extensions.org¹⁵, PRONOM¹⁶) navedejo tudi programske opreme, ki ta format podpirajo. Včasih pa ne vemo, v katerem formatu so dokumenti, ker ne poznamo priponke ali pa je ta napačna. Tudi v tem primeru so na voljo posebna programska orodja, npr. DROID¹⁷, JHOVE¹⁸, FileMerlin¹⁹, XENA²⁰, CheckFileType.com²¹ in veraPDF (PDF/A validation)²².

¹⁵ Dostopno na <https://www.file-extensions.org/> (24. 2. 2018).

¹⁶ Dostopno na naslovu <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx> (24. 2. 2018).

¹⁷ Dostopno na naslovu <http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/> (24. 2. 2018).

¹⁸ Dostopno na naslovu <http://jhove.sourceforge.net/> (24. 2. 2018).

¹⁹ Dostopno na <http://www.file-convert.com/fmn.htm> (24. 2. 2018).

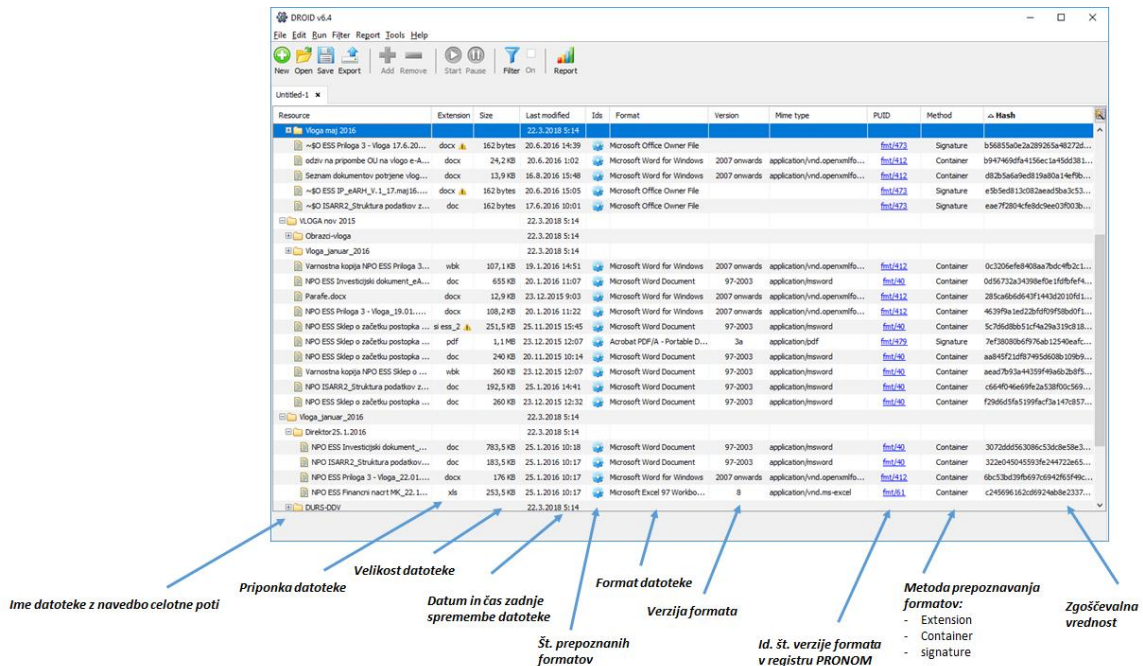
²⁰ Dostopno na <http://xena.sourceforge.net/> (24. 2. 2018).

²¹ Dostopno na <http://www.checkfiletype.com/> (24. 2. 2018).

²² Dostopno na <http://openpreservation.org/about/projects/verapdf/> (24. 2. 2018).

DROID (Digital Record Object Identification)

Programsko orodje DROID (zadnja različica je 6.4) je razvil The National Archives. Z njegovo pomočjo lahko avtomatsko identificiramo format datoteke še pred dostopom do nje oz. njenim odpiranjem, torej ne glede na priponko v njenem imenu. Povezuje se z registrom PRONOM, trenutno je sposoben prepoznati preko 1400 formatov, se pa še naprej razvija. Poleg prepoznavanja formatov pokaže programsko orodje DROID še druge podatke o datoteki, npr. velikost datoteke, zadnji spremenjeni datum in pot do datoteke (slika 5). Te informacije so predstavljene v profilu, ki ga je mogoče analizirati na zaslonu v grafičnem uporabniškem vmesniku DROID (GUI) s filtriranjem ali izvoziti v datoteko CSV.



Slika 5: Programsko orodje DROID identificira formate ter prikaže še druge podatke o posameznih e-zapisih

Priponka datoteke

Identifikacija na podlagi priponke (angl. *extension*) pomeni, da je bil format identificiran zgolj na podlagi imena priponke datoteke. Takšna identifikacija ni nujno zanesljiva, ker lahko datoteke poimenujemo na kakršenkoli način vključno z določitvijo priponke. Npr., dokument test.docx je ustvarjen z MS Word, vendar ga lahko kasneje preimenujemo v test.pdf in ta dokument je še vedno v formatu MS Word, a ga po preimenovanju ne bomo mogli več odpreti.

Format in njegova verzija

DROID na podlagi notranjih lastnosti dokumentov ugotovi format in njegovo različico posamezne datoteke vključno s številko PUID.

PUID (Persistent Unique Identifier)

PUID predstavlja identifikacijsko številko formata v registru PRONOM, kjer najdemo podroben opis različice formata vključno s seznamom programskih orodij, ki prepoznajo format.

Metoda prepoznavanja formata

Identifikacija na podlagi *podpisa* (angl. *signature*) pomeni, da je bil format identificiran z iskanjem vzorcev podpisa v datoteki, za katerega je znano, da se pojavlja v določenih formatih datotek in različicah. Ta metoda je precej zanesljiva, saj je malo verjetno, da se bo slučajno zgodilo, da ima datoteka vzorec, ki pripada drugemu formatu datoteke, kot je njen.

Identifikacija na podlagi *vsebnika* (angl. *container*) pomeni, da je bil format datoteke identificiran z iskanjem vdelanih datotek (po možnosti s svojimi podpisi) znotraj glavne datoteke. Na primer, datoteke ustvarjene v programu Microsoft Office 2010 so dejansko datoteke zip, ki vsebujejo datoteke xml, slike ali druge vire, uporabljene v dokumentu. Identifikacija na podlagi vsebnika bi identificirala glavno datoteko kot datoteko Microsoft Office 2010, ne pa kot datoteko zip. Ta metoda je zelo zanesljiva, saj prepozna širšo vrsto vsebnika (npr. zip), nato identificira datoteke znotraj datoteke zip na podlagi njihovih notranjih lastnosti.

HASH (zgoščevalna vrednost)

Zgoščevalna vrednost (angl. *Hash*) predstavlja dolge številke, izračunane iz vsebine datoteke. DROID lahko ustvari zgoščevalne vrednosti MD5, SHA1 ali SHA256. Na podlagi primerjanja zgoščevalnih vrednosti posamezne datoteke v različnih časovnih obdobjih lahko ugotovljamo, ali je datoteka še vedno enaka originalu, saj se namreč najmanjša sprememba v datoteki (npr. dodana ena pika v besedilu datoteke) odraža v spremenjeni zgoščevalni vrednosti. Vendar pa izračunavanje zgoščevalne vrednosti v orodju DROID v prvi vrsti ni namenjeno iskanju podvojenih datotek, kar bomo prikazali v nadaljevanju.

7.2 Avtomatizirano določanje elektronskih arhivskih zapisov

7.2.1 Lociranje in označitev zapisov, ki vključujejo izbrane pojme oziroma tipskih zapisov

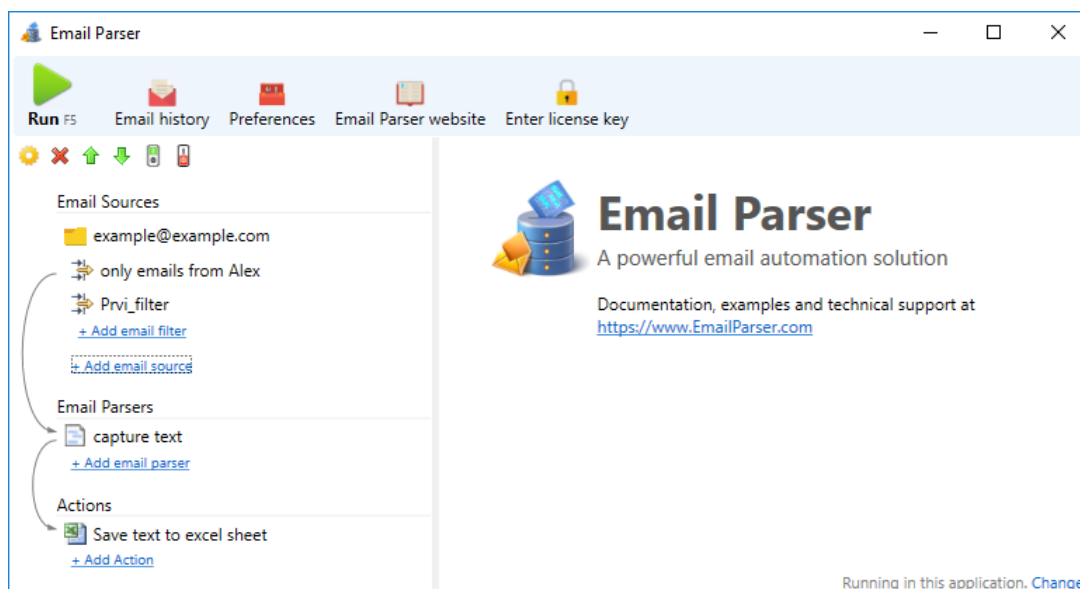
Z uporabo različnih programskih orodij lahko avtomatiziramo označitve, kateri zapisi imajo lastnosti arhivskih, kar je mogoče izvesti na podlagi vnaprej opredeljenega nabora posebnosti oz. značilnosti. S programskimi orodji lahko lociramo in nato označimo vse zapise, ki vsebujejo določeno ime osebe, kraj, dogodek (njegovo omembo), določen datum ali posebne vsebine zapisov (npr. oznake tajnosti, opredelitev posameznih aktov, kot so sodbe, odločbe, sklepi, tipološki zapisi). V tem primeru govorimo o avtomatizirani (strojni) označitvi e-zapisov, ki imajo arhivsko vrednost, torej o *strojni izvedbi dela procesa vrednotenja (valorizacije)*. To lahko podpremo s programskimi orodji, s katerimi lahko avtomatizirano generiramo metapodatke v obliki njihove ekstrakcije (angl. *extraction*) iz različnih vrst e-zapisov.

V ta namen je razvitih že kar nekaj programskih orodij, mi jih izpostavljam samo nekaj:

- **Apache Tika**²³ (razvijalec The Apache Software Foundation). Orodje za analizo vsebine. Odkriva in izvleče metapodatke in besedilo iz več kot tisoč različnih vrst datotek, poleg tega pa zagotavlja tudi knjižnico Java, ima strežniške in ukazne vrstice, primerne za uporabo iz drugih programskih jezikov.

²³ Zadnja različica orodja *The Apache Tika*[™] 1.17 je objavljena decembra 2017 in je dostopna na naslovu <https://tika.apache.org/> (23. 2. 2018).

- **Metadata Extraction Tool**²⁴ (razvijalec Nacionalna knjižnica Nove Zelandije). Orodje avtomatično izloči (izvleče) metapodatke, ki so relevantni za dolgoročno hrambo e-zapisov, skladno z za te namene pripravljeno metapodatkovno shemo ter jih shrani v formatu XML.
- **AVI MetaEdit**²⁵ (razvijalec National Archives and Records Administration): omogoča urejanje metapodatkov za datoteke AVI. Orodje lahko uporabimo za vdelavo, urejanje, uvoz in izvoz metapodatkov.
- **ExSite9**²⁶ (razvijalec Intersect Australia za Australian National Data Service): namizna aplikacija, z njo enostavno in hitro označimo svoje podatkovne datoteke z opisnimi metapodatki ter nato podatkovne datoteke in z njimi povezane metapodatke zapakiramo ter pripravimo za predložitev v e-repozitorij. Omogoča tudi strukturno organizacijo omenjenih datotek pri njihovem premikanju po fizičnih lokacijah; s tem omogoča pravilno organiziranje datotek in metapodatkov, pripravljenih za pakiranje.
- **Email Parser**²⁷ (razvijalec Rockefeller Archive Center in Smithsonian Archives): programsko orodje za obdelavo dohodnih e-sporočil (slika 6).



Slika 6: Osnovni meni orodja Email Parser

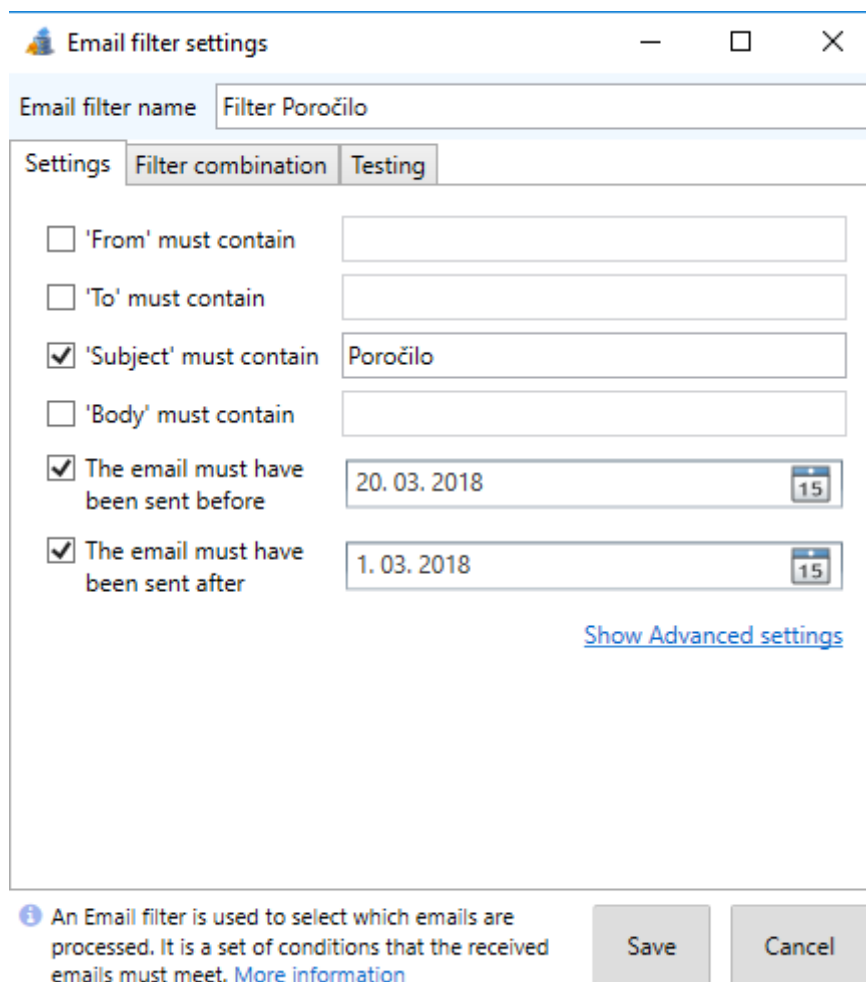
Orodje omogoča ekstrakcijo podatkov iz glave (angl. header) in telesa (angl. body) e-sporočil in s tem pretvorbo nestrukturirane e-pošte v strukturirane podatke, ki jih je mogoče enostavno obdelati.

²⁴ Dostopno na <https://sourceforge.net/projects/meta-extractor/> (24. 2. 2018).

²⁵ Dostopno na <https://mediaarea.net/AVIMetaEdit> (25. 2. 2018).

²⁶ Dostopno na <https://github.com/IntersectAustralia/exsite9> (25. 2. 2018).

²⁷ Testna različica je dostopna na <https://www.emailparser.com/> (23. 3. 2018).

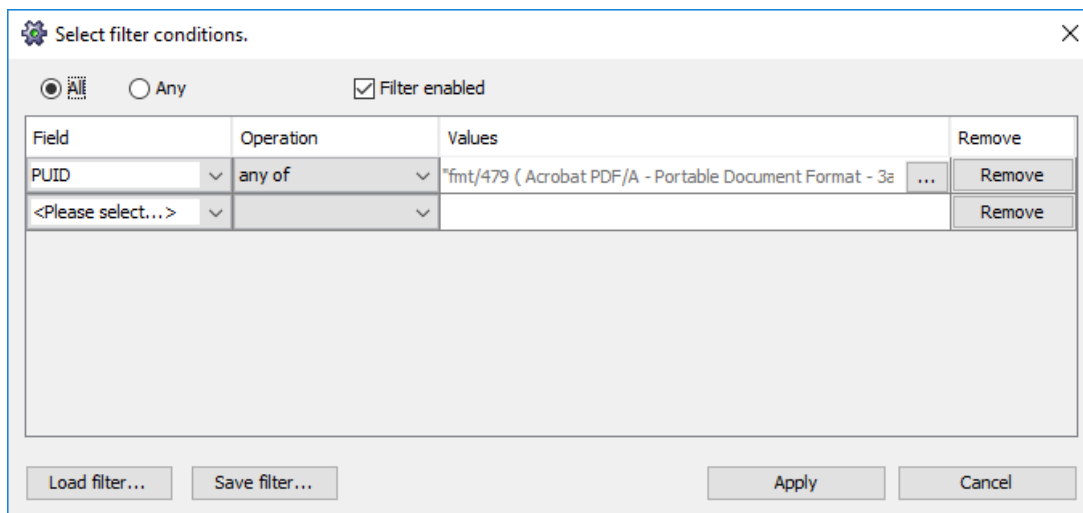


Slika 7: Možnosti filtriranja e-sporočil

Razčleni lahko tudi podatke iz prilog e-pošte (PDF, Excel). S tem dosežemo avtomatizacijo obdelave e-sporočil. Orodje ima možnost različnih nastavitev za iskanje specifičnih delov besedila e-sporočil (angl. *email parsing*), omogoča npr. filtriranje e-pošte glede na pošiljatelja, predmet, vsebino sporočila, datum itd. ter omogoča kombiniranje različnih pravil filtriranja e-pošte za bolj zapleteno izbiranje podatkov iz e-pošte (slika 7). Na tak način lahko izberemo zgolj tista e-sporočila, ki ustrezajo vnaprej določenim kriterijem, pridobljene podatke lahko shranimo tudi v Excel. Omogoča tudi beleženje zgodovine (sledenje obdelavi prejete e-pošte).

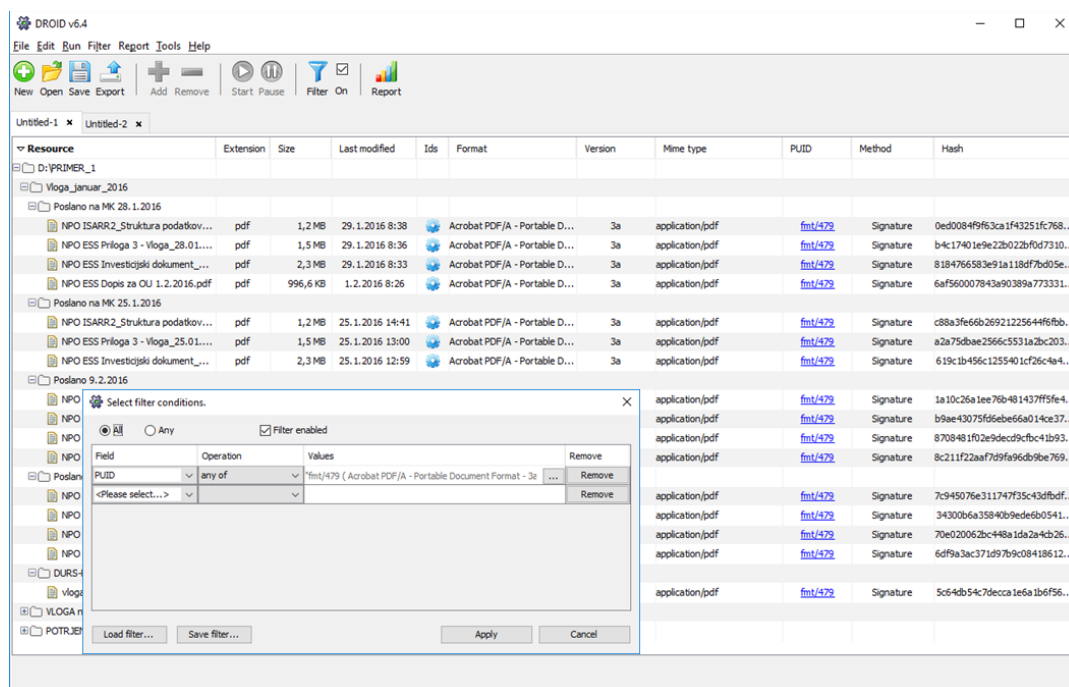
- **Dependency Discovery Tool**²⁸ (razvijalec School of Engineering in Computer Science at Victoria University, Wellington, Archives New Zealand): orodje za odkrivanje povezav med datotekami .doc, .xls in .ppt, skuša najti vse dokumente ali datoteke, ki so povezani z nekim dokumentom, kot izhodni dokument lahko ustvari .txt, .xml ali .csv. Napisano je v Javi z uporabo knjižnic Apache POI (<http://poi.apache.org>).
- Z že prej omenjenim orodjem **DROID** lahko vnaprej določimo, katere datoteke nas v nadaljevanju zanimajo, tako da v Filtru določimo kriterije izbora, npr. ime datoteke, priloga ali ime formata, številka PUID (slika 8).

²⁸ Dostopno na <https://sourceforge.net/projects/officeddt/> (23. 3. 2018).



Slika 8: Možnost nastavitve kriterijev izbora datotek v orodju DROID

Kot rezultat dobimo ožji nabor datotek glede na izbrane kriterije. Če npr. želimo, da iz nabora e-zapisov dobimo samo datoteke, ki so v formatu PDF/A-3 PUID= fmt/479), bomo dobili seznam samo tistih datotek, ki ustrezajo temu pogoju (slika 9).



Slika 9: DROID pripravi seznam datotek glede na vnaprej določene kriterije (Filter)

7.3 Tehnično (strojno) vrednotenje elektronskih zapisov

73.1. Deduplikacija

Napredne tehnologije uporabnikom omogočajo, da shranijo vsako e-sporočilo in praktično katerokoli drugo vrsto e-zapisov, ki jih ustvarijo, in mnogi to tudi storijo – in to že vrsto let. To povzroča kopičenje velike količine e-zapisov, tudi njihovo dupliciranje²⁹, s tem pa tudi nepreglednost nad njimi. E-zapisi imajo lahko več kopij v več različicah, shranjeni so lahko na več različnih lokacijah, po možnosti pod različnimi imeni in v različnih formatih. Kot primer vzemimo slike in videoposnetke. Fotografije, posnete na mobilnih napravah, se lahko na primer samodejno shranijo v iCloud³⁰, morda se kopirajo v Facebook/Flickr ali Instagram, kasneje pa se prenesejo v namizni računalnik, kamor se samodejno kopirajo v Google Foto. Vendar je to le eden od primerov izmed množice drugih. Zaradi napačne izbire oz. določitve, kaj je izvirnik, lahko pride do tega, da ohranimo na primer fotografije, ki so slabe kakovosti. Nalaganje fotografij in videoposnetkov na zunanja spletna mesta namreč pogosto povzroči izgubo kakovosti, če so shranjene v stisnjenem formatu. Čeprav se primeri fotografij in videoposnetkov večinoma zgodijo v privatnem življenju, pa se lahko pojavijo tudi v institucionalnih okoljih.

Podoben izziv najdemo tudi v institucionalnih zbirkah. Avtorjev e-zapisov v okviru ene raziskave je lahko več, hranijo se lahko v različnih repozitorijih v različnih verzijah (npr. delovna verzija, objavljena verzija). Verzije so lahko nedosledno označene v različnih e-repozitorijih, obstaja pa tudi tveganje, da spremembe na eni različici e-zapisa ne bodo prenesene v vse druge e-repozitorije. Tudi to povečuje tveganje izgube podatkov in spodkopava avtentičnost (verodostojnost) e-zapisov.

Ob obilici e-zapisov, ki se pojavljajo v dvojnikih, trojnikih oziroma v večkratnih različicah, je zato smiselno razmišljati o t. i. *tehničnem vrednotenju* (angl. *technical appraisal*), v določenih primerih pa tudi o *vsebinskem vrednotenju* (če se izkaže, da e-zapisi niso isti) ter nato *tehničnem odbiranju* (angl. *technical selection*). Postopek odpravljanja podvojenih zapisov je v računalništvu poznan kot proces deduplikacije (angl. *deduplication*) in se običajno uporablja za sprostitev zasedenih pomnilnih kapacitet ter izboljšanja zmogljivosti diskovnih sistemov. Deduplikacijo lahko izvajamo na nivoju datotek tako, da algoritmi med seboj primerjajo datoteke, ali pa na nivoju podatkovnih blokov, kjer algoritmi med seboj primerjajo bloke podatkov. Če so primerjane datoteke med seboj enake, se ohrani le en unikatni zapis te datoteke.

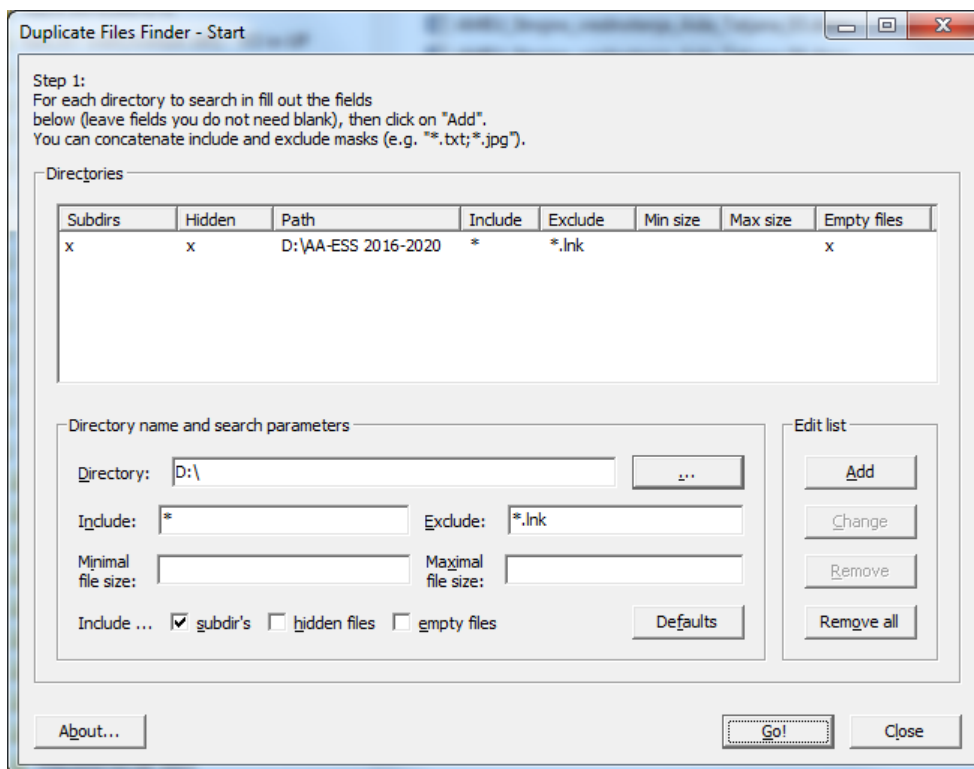
Za deduplikacijo e-zapisov, kot načina tehničnega vrednotenja in odbiranja, lahko uporabimo različna programska orodja, kot primer najprej navajamo **Duplicate Files Finder**³¹ (razvijalec Matthias Böhm). Program omogoča iskanje dvojnikov znotraj dane mape. Iskanje lahko omejimo z določenimi parametri. Iščemo lahko npr. tudi po podmapah, skritih datotekah in iščemo prazne datoteke. Vključimo lahko več različnih lokacij. Iskanje lahko dodatno omejimo tako, da iščemo samo dvojnike pri določenih končnicah datotek (npr. .docx, .doc, .xlsx, .pptx, .pdf) ter izpustimo določene tipe datotek,

²⁹ Raziskava podjetja FindLaw navaja, da je v industriji povprečno 21 % podvojenih e-dokumentov (glej: <http://technology.findlaw.com/ediscovery-guide/processing-metrics.html> (25. 2. 2018)).

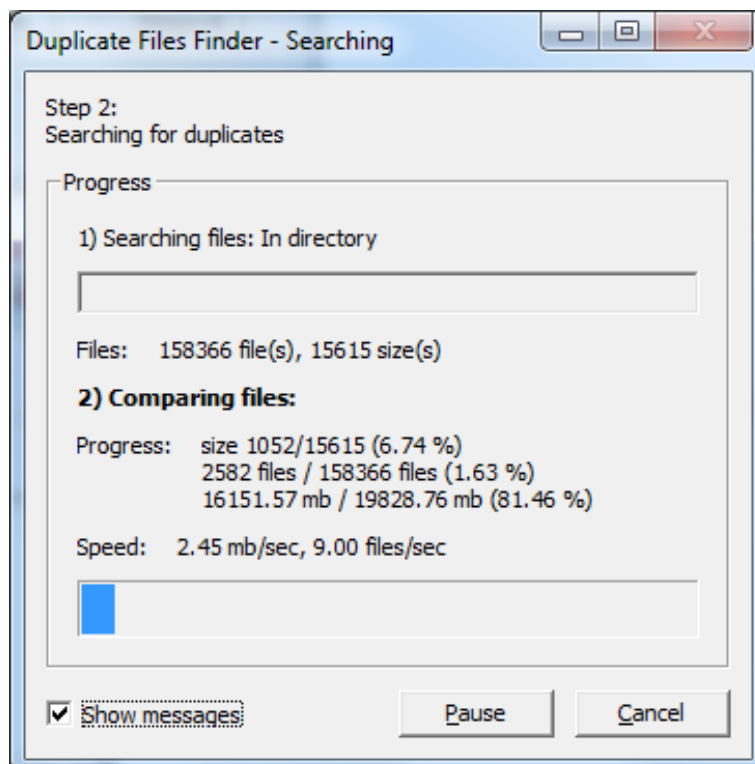
³⁰ iCloud je Appleov storitev, ki samodejno povezuje aplikacije in naprave z operacijskih sistemom iOS 5, deloma pa ga lahko uporabljamo tudi na Mac in Windows računalnikih. S prijavo na storitev iCloud dobimo brezplačen me.com e-poštni račun, iCloud pa poskrbi za potisk nove pošte prav na vse naše iCloudove naprave. Hkrati iCloud poskrbi tudi za sinhronizacijo koledarjev in kontaktov, tako na mobilnih napravah kot na Mac-u ali PC-ju. iCloud poskrbi tudi za to, da imajo vse naše iNaprave iste aplikacije, knjige (iz aplikacije iBooks) in da se v oblaku redno shranjuje varnostna kopija naših podatkov in nastavitve za primer, da gre kaj narobe. iCloud zna poleg vsega tega med našimi napravami samodejno prenašati tudi dokumente.

³¹ Dostopno na naslovu <http://doubles.sourceforge.net/> (24. 2. 2018).

npr. pri slikah Thumbnail. Dodatno pa lahko iskanje omejimo po velikosti tako, da določimo minimalno ter maksimalno velikost datoteke (sliki 10 in 11).

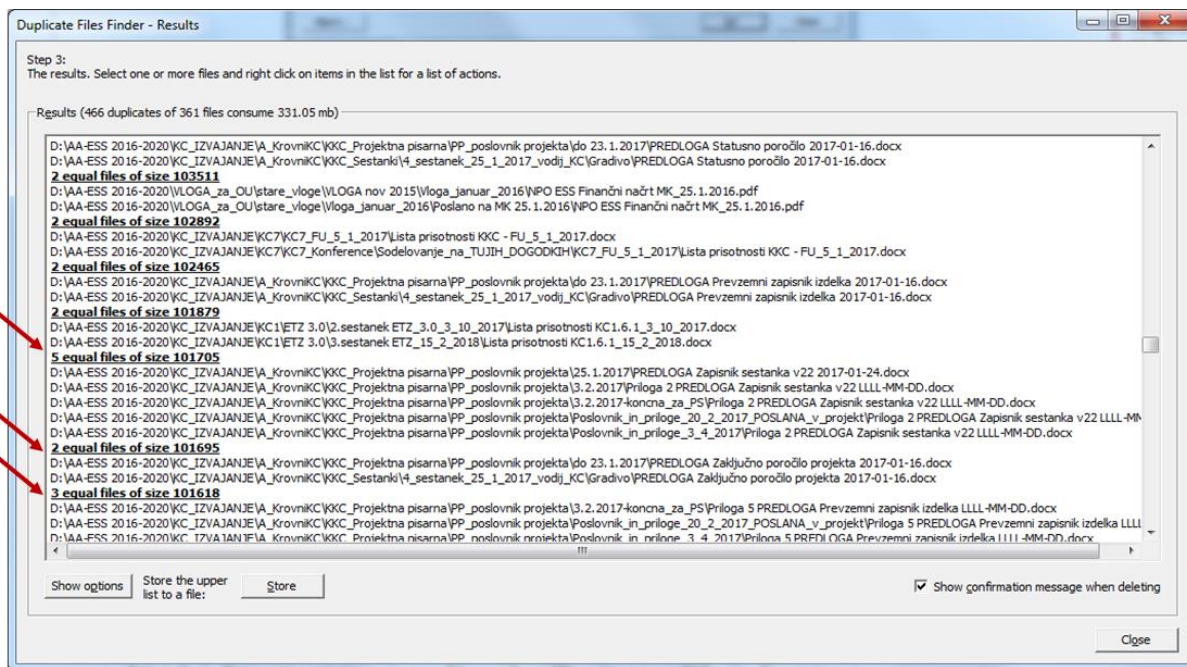


Slika 10: Program Duplicate Files Finder – ekranska slika z možnostjo nastavitvev



Slika 11: Program Duplicate Files Finder primerja datoteke in išče dvojnike

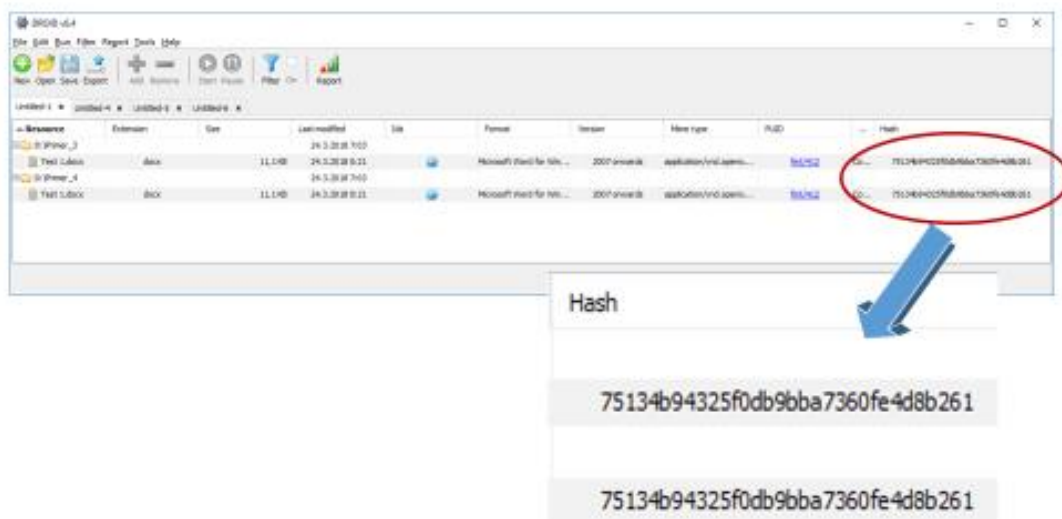
Program išče po izbranih mapah in parametrih ter ustvari seznam datotek, ki jih je našel v eni ali več kopijah. Na sliki 12 vidimo, da je program določene datoteke našel v dveh, treh in celo v petih kopijah.



Slika 12: Rezultat iskanja dvojnikov s programom Duplicate Files Finder

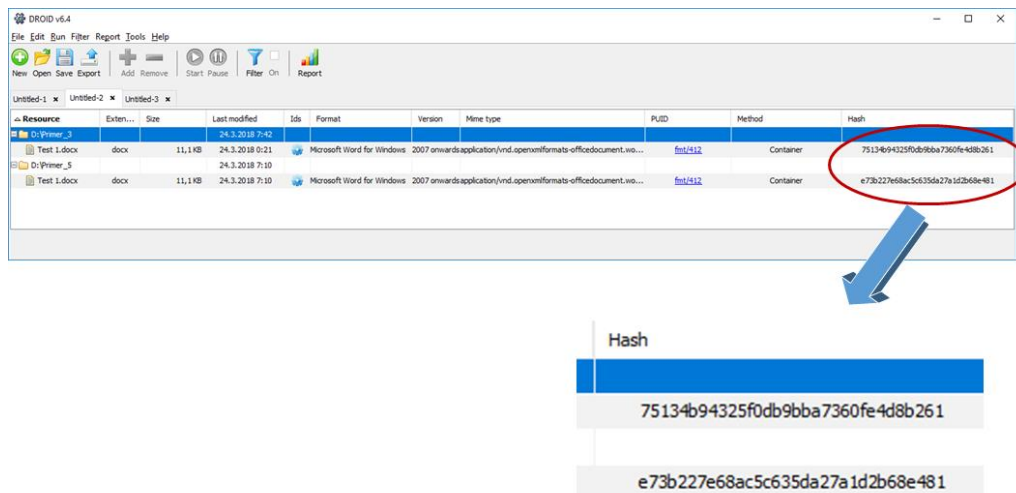
Iskanje dvojnikov si je skoraj nemogoče predstavljati, brez da bi si pri tem pomagali s programskimi orodji. Če pogledamo seznam dvojnih datotek, vidimo, da imajo določene datoteke res enaka imena, vendar na podlagi tega podatka še ni mogoče sklepati, da gre za res isto datoteko. Seveda pa program Duplicate Files Finder ni iskal dvojnikov zgolj na podlagi imena datoteke, poleg vsebine je primerjal še druge lastnosti datotek.

Ena izmed metod hitrega iskanja dvojnikov temelji tudi na izračunu in primerjanju zgoščevalne vrednosti (angl. *hash*) datotek. Če imata dve datoteki enako zgoščevalno vrednost, potem imata verjetno enako vsebino (*verjetnost, da imata dve poljubno izbrani datoteki enako zgoščevalno vrednost, izračunano npr. z algoritmom MD5, je 2^{-128} , kar je zelo, zelo nizko*). Z orodjem DROID lahko generiramo zgoščevalne vrednosti (slika 13), vendar pa samo orodje DROID ne išče dvojnikov. Seznam datotek z izračunanimi zgoščevalnimi vrednostmi lahko shranimo npr. v Excel ali CSV ter nato znotraj tega iščemo datoteke z enako zgoščevalno vrednostjo.



Slika 13: Z orodjem DROID lahko enostavno izračunamo zgoščevalno vrednost za vsako datoteko

Zgoščevalna vrednost se izračuna na podlagi vsebine datoteke in pripadajočih metapodatkov. Na sliki 13 vidimo primer dveh datotek v različnih direktorijih (mapah) z enakim imenom, enako vsebino, enakim datumom in časom nastanka. V naslednjem koraku smo eni izmed obeh datotek dodali naslov datoteke kot metapodatek (možnost dodajanja lastnosti datoteke znotraj urejevalnika MS Word). Ime in vsebina datoteke je ostala ista, razlika je le v dodanem metapodatku in posledično tudi datumu zadnje spremembe datoteke. Ko smo ponovno izračunali zgoščevalni vrednosti obeh datotek, sta bili ti različni (slika 14).



Slika 14: Izračun zgoščevalne vrednosti po dodanih metapodatkih v eno izmed primerjanih datotek

Na podlagi tega lahko trdimo, da gre v primeru, ko so zgoščevalne vrednosti več datotek popolnoma enake, za čiste dvojnike. Na tej točki bi se lahko odločili za ohranjanje dvojnikov le zato, ker se ti nahajajo na različnih lokacijah (npr. v različnih zadevah) in je ta podatek ključen pri nadaljnjem obravnavanju z vidika arhivistike (npr. zaradi kontekstualnih podatkov). Kljub temu pa lahko to metodo uporabimo pri odstranjevanju dvojnikov npr. takrat, ko zagotovo vemo, da so dvojne datoteke na mestih, kjer imamo varnostne kopije, npr. mape, ki vsebujejo imena "backup", "temp" in "old", saj uporabniki začasne kopije pogosto imenujejo mape ali datoteke s temi besedami.

Vendar pa predstavljene metode lahko dajo le navedbo, da imamo podvojene datoteke in se bo treba odločiti, kako ravnati z njimi. Zato bo še vedno potrebna visoka stopnja ročnega pregleda, da se bomo prepričali, da imamo dejansko dvojnike datotek, ki jih ne potrebujemo in se bomo morali odločiti, katere datoteke so lahko varno odstranjene. Ker obstajajo tudi določena tveganja pri brisanju datotek, je treba pred brisanjem dvojnikov upoštevati še druge vidike, npr.:

- Vsi uporabniki ne smejo imeti dostopa do vseh kopij podvojenih datotek.
- Podvojene datoteke imajo lahko različne metapodatke. Na primer, lokacija datotek v določeni strukturi ima lahko bistven kontekstualni pomen.
- Podvojena datoteka lahko zagotovi kontekst okoliških podatkov, ki so bili shranjeni zraven te datoteke. Izguba tega konteksta bi lahko povzročila, da bi bil pomen okoliških datotek nejasen ali nerazumljiv.

Nekatere od tveganj, povezanih z izgubo konteksta, lahko ublažimo tako, da pustimo bližnjice (angl. *shortcuts*) v »glavni datoteki«, ko izbrišemo njene dvojnike.

7.3.2 Preverjanje avtentičnosti in celovitosti e-zapisov

E-zapisi morajo biti celoten življenjski cikel avtentični in celoviti. Za to pa morajo biti izpolnjeni določeni tehnični pogoji. Za oceno avtentičnosti in celovitosti lahko uporabimo različne validacijske metode, ki jih izvajamo s pomočjo programskih orodij, npr. preverjanje zgoščevalne vrednosti ali elektronskih podpisov; vse to pa sodi v *tehnično (strojno) vrednotenje*.

Omenili smo že orodje DROID, ki na podlagi notranjih lastnosti datotek izračuna zgoščevalno vrednost, ki jo vedno lahko kasneje preverjamo in ugotavljamo nespremenljivost datotek. Drugo takšno orodje je npr. **ACE**³² (razvijalec University of Maryland Institute for Advanced Computer Studies), ki predstavlja skupek orodij, razvitih za arhive in knjižnice, da preverijo na podlagi močnih kriptografskih tehnik, da je gradivo, ki ga hranijo, celovito in zaupanja vredno.

7.3.3 Programska orodja kot pomoč pri strojnem vrednotenju e-zapisov

Na spletnih straneh je, poleg že prej omenjenih, mogoče najti tudi še vrsto drugih programskih orodij, s pomočjo katerih naj bi bilo mogoče izvajati strojno vrednotenje e-zapisov. V nadaljevanju jih nekaj navajamo, vendar jih v okviru naše raziskave zaenkrat še nismo preizkusili:

³² Dostopno na <http://coptr.digipres.org/ACE> (*Audit Control Environment*) (24. 2. 2018).

Tabela 1: Seznam programskih orodij za pomoč pri vrednotenju e-zapisov

Orodje	Razvijalec	Dostopno na	Področje funkcionalnosti
<i>ePADD</i> (Email: Process, Appraise, Discover, Deliver)	Project MUSE pri Stanford University Computer Science Department	http://epadd.stanford.edu/epadd/collections	E-pošta Metapodatki Vizualizacija Vrednotenje
<i>Pagelyzer</i>	projekt SCAPE	https://github.com/openpreserve/pagelyzer	Primerjava spletnih strani Vrednotenje
<i>DELOLD</i>	Alex Issakoo	https://sourceforge.net/projects/delold/	Čiščenje datotek Upravljanje datotek Upravljanje map v skupni rabi
<i>NARA File Analyzer and Metadata Harvesting Tool</i>	National Archives and Records Administration	https://github.com/usnationalarchives/File-Analyzer https://narations.blogs.archives.gov/2011/10/18/national-archives-digitization-tools-now-on-github	Iskanje napak (checksum) pri prenosu ali pretvorbi datotek
<i>BitCurator</i>	School of Information and Library Science at the University of North Carolina, Chapel Hill (SILS) and the Maryland Institute for Technology in the Humanities (MITH)	http://bitcurator.net/	Digitalna forenzika Analiza datotečnega sistema Upravljanje metapodatkov Iskanje zaupnih informacij (Personally Identifiable Information)
<i>BWF MetaEdit</i>	Federal Agencies Digitization Guidelines Initiative	http://bwfmetaedit.sourceforge.net/	Upravljanje metapodatkov Audiovizualne datoteke
<i>browser-shots</i>	Internet Memory SCAPE Project	https://github.com/sbarton/browser-shot-tool-mapred	Vrednotenje Ohranjanje
<i>EXIF to DC XML Normalizer</i>	Open Planets Foundation	https://github.com/opflabs/SPRUCE/tree/master/exif_to_dc	Upravljanje metapodatkov
<i>GNU Diffutils</i>	Free Software Foundation	http://www.gnu.org/software/diffutils/	Iskanje razlik med datotekami
<i>Matchbox</i>	SCAPE Project	http://openpreservation.org/matchbox/	Iskanje duplikatov Slikovno gradivo
<i>Muse</i>	The Stanford Mobile and Social Computing Research Group	https://mobisocial.stanford.edu/muse/	E-pošta Vrednotenje
<i>veraPDF</i>	PDF Association	http://verapdf.org/home/#about	PDF validator

7.4 Strojno odbiranje elektronskih arhivskih zapisov

Po fazi vrednotenja sledi faza odbiranja e-zapisov, ki jo prav tako lahko izvedemo s pomočjo programskih orodij. Seveda je izjemno pomembno, v katerem okolju smo locirali e-zapise, ki jih vrednotimo in odbiramo za dolgoročno hrambo oz. kot arhivsko gradivo. Dober primer, kjer lahko uporabimo strojno odbiranje, najdemo v okviru informacijskih sistemov za upravljanje z dokumenti (ISUD). Podlago za odbiranje predstavlja klasifikacijski načrt.

Uporabo klasifikacijskega načrta na Slovenskem v državni in lokalni upravi datiramo od 18. stoletja dalje. V obdobju od prve svetovne vojne do leta 1955 se klasificiranje dokumentov ni uporabljalo, potem pa so v takratnih republiških, pokrajinskih in občinskih upravnih organih ponovno začeli vpeljevati klasificiranje in arhiviranje dokumentov, torej razvrščanje oziroma strukturiranje posameznih enot gradiva po pristojnostih, funkcijah, nalogah in dejavnosti (Žumer, 2013). Klasifikacija dokumentov namreč omogoča vpogled v vsebino, preglednost, dostop in določanje rokov hrambe.³³ Dodatno pa klasifikacija oziroma klasifikacijski načrt omogoča pregled rokov hrambe in določila o arhivskem gradivu, ki ima trajen pomen za zgodovino, druge znanosti in trajni pravni interes fizičnih in pravnih oseb.³⁴ Rok hrambe, ki so v klasifikacijskem načrtu, morajo biti usklajeni s predpisi in potrebami poslovanja, omogočajo izločanje in uničenje nepotrebne dokumentacije, ki so ji potekli roki hrambe in ki nima več pomena za tekoče poslovanje, po predpisanem postopku (ZVDAGA, 13.–16. čl.). Določila o arhivskem gradivu, ki jih v klasifikacijski načrt vnesejo pristojni državni arhivi, pa omogočajo odbiranje arhivskega gradiva in s tem njegovo izročanje pristojnim arhivom, v kolikor gre za javno arhivsko gradivo.

Zahteva po uvedbi in uporabi klasifikacijskega načrta z roki hrambe kot osnove za razvrščanje dokumentacije na podlagi funkcij, nalog, pristojnosti na vseh področjih je bila zajeta tudi v standardu ISO 15489-1:2001. V Sloveniji je omenjeni standard ISO 15489³⁵ upoštevan v skoraj vseh predpisih, ki se nanašajo na upravljanje z dokumenti, že od samega sprejetja, kljub dejstvu, da je bil šele leta 2005 sprejet kot slovenski nacionalni standard.

Zasnova in uporaba klasifikacijskega načrta podajata realno in uporabno podlago za *avtomatiziran proces odbiranja*. To potrjujejo tudi zahteve v III. delu Enotnih tehnoloških zahtev³⁶ (2013), ki npr. določajo:

- ISUD³⁷ mora biti sposoben podpirati načrt razvrščanja gradiva (klasifikacijski načrt), ki lahko predstavlja zadeve in dokumente, hierarhično organizirane v razrede (ETZ III/3.3.1.1).
- ISUD mora podpirati najmanj tri ravni v hierarhiji načrta razvrščanja gradiva (ETZ III/3.3.1.3).
- Ko ISUD uvozi celoten načrt razvrščanja gradiva ali njegov del, mora omogočiti uvoz povezanih metapodatkov o rokih hrambe in revizijskih sledi, če obstajajo (ETZ III/3.3.1.5).

³³ Osnove klasifikacije je obdelal Schellenberg v *Modern Archives, Principles & Techniques*.

³⁴ Arhivsko gradivo je v Sloveniji definirano v 2. členu ZVDAGA (2014).

³⁵ Leta 2016 je sprejeta posodobljena različica standarda ISO 15489-1:2001 (ISO: <https://www.iso.org/obp/ui/#iso:std:iso:15489:-1:ed-2:v1:en>)

³⁶ Enotne tehnološke zahteve so dokument, ki ga ZVDAGA iz leta 2006 neposredno določa v 91. členu, trenutna različica 2.1 je še v veljavi. Leta 2014 novelirani ZVDAGA pa je ta dokument »prekvalificiral« v Pravilnik o enotnih tehnoloških zahtevah v 17. členu in je v pripravi v Arhivu RS.

³⁷ ISUD - Informacijski sistem za upravljanje z dokumenti.

- ISUD mora v okviru metapodatkov razreda ali zadeve shraniti datum izdelave in zaprtja razreda ali zadeve (ETZ III/3.3.2.5).

Nato pa so opredeljene še zahteve za ISUD, ki mora podpirati postopek odbiranja in izločanja (poglavje 3.5), npr.:

- ISUD ne sme omejevati števila rokov hrambe, odbiranja in izločanja (ETZ III/3.5.1.1).
- ISUD mora ohranjati nespremenljivo zgodovino sprememb in izbrisov (revizijsko sled), ki se nanašajo na roke hrambe, odbiranja in izločanja, vključno z datumom spremembe ali izbrisa in uporabnikom, ki je vnesel spremembo (ETZ III/3.5.1.2).
- ISUD mora zagotoviti, da se vsaka sprememba v roku hrambe, odbiranja in izločanja nemudoma uporabi za vse entitete, ki jim je bil rok hrambe, odbiranja in izločanja dodeljen (ETZ III/3.5.1.3).
- ISUD mora biti sposoben uvažati in izvažati roke hrambe, odbiranja in izločanja (ETZ III/3.5.1.4).
- Odbiranje in izločanje vsakega dokumenta morata biti vodeni z rokom/i hrambe, odbiranja in izločanja, povezanim(i) z vrsto razreda, zadeve, in vrste dokumenta, ki jim dokumenti pripadajo, ter s katerim koli veljavnim zadržanjem uničenja (ETZ III/3.5.1.6).
- ISUD mora omogočati vsaj te ukrepe v zvezi z odbiranjem in izločanjem za vsak rok hrambe, odbiranja in izločanja (ETZ III/3.5.1.7):
 - trajna hramba;
 - pripraviti za pregled;
 - uničiti po potrditvi administratorske vloge v sistemu ISUD;
 - prenos v arhiv ali drugo skladišče.

V kolikor torej imamo klasifikacijski načrt, uvožen v ISUD, lahko že na podlagi podatkov iz njega strojno odberemo zapise, ki ustrezajo določenim pogojem oz. metapodatkom (npr. oznaka A pomeni, da so zapisi arhivski, določimo lahko datumski okvir, pomeni, da se odberejo zapisi, ki imajo datum znotraj tega okvira).

8. Zaključek

Raziskovalno področje, ki je predstavljeno v prispevku, izhaja iz potrebe po vrednotenju in odbiranju e-zapisov, ki jih je ob vedno bolj intenzivnem e-poslovanju tudi vedno več. Pregled literature in dobrih praks, dostopnih na spletu, so pokazale, da so pri iskanju rešitev na tem področju že vrsto let aktivni nacionalni arhivi po svetu. To je nekako razumljivo, saj je na eni strani njihova primarna skrb, da ohranjajo kulturno dediščino svojega naroda, ne glede na to, v kakšni obliki je ta, po drugi strani pa je očitno eksponentno naraščanje e-zapisov pri ustvarjalcih arhivskega gradiva. Dokler so bili zapisi samo na fizičnih nosilcih, praviloma papirju, so bili arhivisti popolnoma suvereni pri njihovem obvladovanju, ki se začne z vrednotenjem in odbiranjem ter nadaljuje z dolgoročnim ohranjanjem in nadaljnjo uporabo. S pojavom e-zapisov pa temu po vsej verjetnosti ne bo več tako. Ocenjujemo, da »ročne« načine vrednotenja in odbiranja, kot se uporabljajo za zapise na fizičnih nosilcih, glede na kompleksnost in eksponentno naraščajočo količino e-zapisov ne bo mogoče enostavno prenesti v elektronsko okolje. E-zapisi, ki so izvorno nastali v elektronskem okolju, ponujajo različne možnosti za uporabo tehnologije pri soočanju z izzivi vrednotenja in odbiranja na način, ki ni bil mogoč

za zapise na fizičnih nosilcih. Uporaba tehnologije za strojno vrednotenje in odbiranje je tudi edina možnost, ki bo v prihodnosti zagotavljala učinkovito delo arhivistov, da bo primerljivo z dosedanjim. Zato bomo raziskave in iskanje rešitev, kako arhivistom lahko pomaga tehnologija pri postopkih vrednotenja in odbiranja e-zapisov, nadaljevali v naslednjem obdobju v okviru projekta e-ARH.si: ESS 2016–2020.

SUMMARY

CAN TECHNOLOGY HELP THE PROCESS OF APPRAISAL AND SELECTION OF E-RECORDS?

Tatjana HAJTNIK, Ph. D.

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

Aida Škoro Babić, M. Sc.

Archives of the Republic of Slovenia, Ljubljana, Slovenia
aida.skoro@gmail.com

The automatization of records appraisal is and will be more and more an increasingly topical issue already due to the increasing number of e-records. In Slovenia and elsewhere in the world, the practice of determination, which record has the archival value has been introduced already for those records on classic carriers (paper) as early as possible in their life cycle - at the time of their creation. Since preserved material is a testimony only if it is preserved in the context, the established rule is that the appraisal requires the scientific-research manner and thus creates the intellectual framework for later selection of archival records. Because of the large number of e-records of different types created on a daily basis (text documents, audio / video, movie, databases, records in document management systems, e-mail, websites, social networks, etc.) it will be impossible to complete the appraisal and selection process in a comprehensive and precise manner and within acceptable time limits. Therefore, it is high time to start thinking about how we can use technology or dedicated software tools to the fullest extent possible. When we begin to think about integrating technology into appraisal and selection procedures, it becomes clear that these procedures can no longer be only in the domain and implementation of archivists, it will be necessary to involve informatics, and in all probability, in certain cases, also lawyers. An overview of the literature and good practices has shown that for many years now national archives around the world have been active in finding solutions in this field. This is somehow understandable, since on the one hand, they are primarily concerned to preserve the cultural heritage of their nation, regardless of the form in which this and, on the other hand, the exponential increase in e-records in the creators of archival material. In this research, there are some answers to the question, which already existed and available tools could be used in individual phase of appraisal and of the archival records selection. Since this research is made in the framework of the project of Slovenian Electronic Archives, this research is going to be improved with new solutions.

Literatura in viri

- Arhiv RS. (2010a).** *Izvedbeni načrt strategije razvoja slovenskega javnega elektronskega arhiva e-ARH.si za obdobje 2010–2015.* Ljubljana: Arhiv Republike Slovenije, 2010, 36 str.
- Arhiv RS. (2010b).** *Strategija razvoja slovenskega javnega elektronskega arhiva e-ARH.si 2010–2015 in Izvedbeni načrt strategije razvoja slovenskega javnega elektronskega arhiva e-ARH.si za obdobje 2010–2015.* Ljubljana: Arhiv Republike Slovenije. http://www.arhiv.gov.si/fileadmin/arhiv.gov.si/pageuploads/zakonodaja/Strategija_e-ARH.si/Strategija_e-ARH_si_2010-2015_v.1.5.pdf [pridobljeno 21. 2. 2018].
- Arhiv RS. (2013).** *Enotne tehnološke zahteve, III. del: Dodatne enotne tehnološke zahteve za ponudnike, strojno in programsko opremo ter storitve, različica 2.1.* http://www.arhiv.gov.si/fileadmin/arhiv.gov.si/pageuploads/E-ARHIVI/ETZ_2_1/ETZ_2.1_III.del_k.pdf [pridobljeno 20. 2. 2018].
- Arhiv RS. (2016).** *Strategija in izvedbeni načrt razvoja slovenskega javnega elektronskega arhiva 2016–2020, različica 1.0.* Ljubljana: Arhiv Republike Slovenije. http://www.arhiv.gov.si/fileadmin/arhiv.gov.si/pageuploads/zakonodaja/Strategija_e-ARH.si/Strategija_e-ARH_si_2016-2020_1.0.pdf [pridobljeno 21. 2. 2018].
- Bearman, D. (1999).** Reality and Chimeras in the Preservation of Electronic Records. *D-Lib Magazine* 5 (4), str. 6–11. Corporation for National Research Initiatives.
- Cook, T. (1992).** Mind Over Matter: Towards a New Theory of Archival Appraisal, V: *The Archival Imagination: Essays in Honour of Hugh A. Taylor*, Ottawa: Association of Canadian Archivists, 1992, str. 38–70.
- Duranti, L. (2001).** The impact of digital technology on archival science. *Archival Science*, Vol 1, str. 39–55.
- Hajtnik, T. (2016).** *Celovit pristop k pretvorbi elektronskih dokumentov v obliko za dolgoročno hrambo : doktorska disertacija.* Maribor.
- Hajtnik, T. (2017a).** Projekt e-ARH.si kot odgovor na izzive dolgoročnega ohranjanja elektronskega arhivskega gradiva. 2. Mednarodna konferenca e-ARH.si, Arhiv RS, Ljubljana, 22. November 2017.
- Hajtnik, T. (2017b).** Vzpostavitev nacionalnega e- arhiva: tukaj je izziv – kako ga lahko rešimo? V: *Atlanti. - Letn. 27, št. 1* (2017), str. 147–164. Trst: Mednarodni inštitut arhivski znanosti Maribor-Trst.
- Harrison, H. P. (1997).** *Audiovisual archives. A practical reader.* Paris: UNESCO.
- IDC. (2014).** *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things.* <https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm> [pridobljeno 6. 9. 2017].
- International Council on Archives (ICA) - Committee on electronic records. (1997).** *Guide for managing electronic records from an archival perspective.* Paris.
- InterPARES1 Project. (2001).** Appraisal Task Force Report, http://www.interpares.org/book/interpares_book_e_part2.pdf [pridobljeno 10. 2. 2018].
- ISO 20652: 2006** Space data and information transfer systems -- Producer-archive interface -- Methodology abstract standard.
- Jenkinson, H. A. (1922).** *Manual of Archival Administration Including the Problems of War Archives and Archive Making.* Oxford: At the Varendon Press.

- Kolsrud, O. (1992).** The evolution of basic appraisal principles – some comparative observations. V: *American Archivist* 55 (1) str. 26–39.
- Levy, D. M. (2000).** *Where's Waldo? Reflections on Copies and Authenticity in a Digital Environment. Authenticity in a Digital Environment.* Washington DC: The Council on Library and Information Resources.
- Lynch, C. (2000).** *Authenticity and Integrity in the Digital Environment: An Exploratory Analysis of the Central Role of Trust. Authenticity in a Digital Environment.* Washington DC: The Council on Library and Information Resources.
- National Archives Administration - National Development Council. (2018).** *Technical Appraisal: Methods of Electronic Record Technical Appraisal.* <https://www.archives.gov.tw/english/Publish.aspx?cnid=982> [pridobljeno 22. 2. 2018].
- Naugler, H. (1984).** *The archival appraisal of machine-readable records: a ramp study with guidelines,* Unesco. <http://unesdoc.unesco.org/images/0006/000635/063501eo.pdf> [pridobljeno 20. 2. 2018].
- Rothenberg, J. (2000).** *Preserving Authentic Digital Information. Authenticity in a Digital Environment.* Washington DC: The Council on Library and Information Resources.
- Schellenberg, T. R. (1956/1984).** *The Appraisal of Modern Public Records.* Washington 1984: Modern Archibves Reader.
- Schellenberg, T. R. (1956/2003).** *Modern Archives. Principles & Techniques.* Chicago: The Society of American Archivists.
- SIST ISO 14721: 2013** Trajno ohranjanje podatkov in sistemi za prenos informacij - Odprti arhivski informacijski sistem (OAIS) - Referenčni model.
- Sklep Vlade RS 81. seje z dne 24. 3. 2016, št. dokumenta 62500-1/2016/5. (2016).** Ljubljana: Vlada RS, 2016.
- Sklep Vlade RS 93. seje z dne 29. 7. 2010, št. dokumenta 62500-2/2010/3. (2010).** Ljubljana: Vlada RS, 2010.
- Society of American Archivists. (2018).** Archival value. *A Glosary of Archival and Records Terminology.* <https://www2.archivists.org/glossary/terms/a/archival-value> [pridobljeno 18. 2. 2018].
- Škoro Babić, A. (2015).** Upravljanje, vrednotenje in dolgoročna hramba e-gradiva: vloga arhivistov in informatikov. V: *Atlanti*, Trst, str. 217–224. Trst: Mednarodni inštitut arhivski znanosti Maribor-Trst.
- Škoro Babić, A. (2017).** Merila za ohranitev arhivskega gradiva. V: *Arhivi v službi človeka – človek v službi arhivov*, str. 146–151. Maribor: Alma Mater Europea.
- 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.
- Žumer, V. (2013).** Klasifikacijski načrt za razvrščanje dokumentacije z roki hranjenja v Sloveniji. V: Gostenčnik, N (ur.) *Tehnični in vsebinski problemi klasičnega in elektronskega arhiviranja, Radenci.* str. 353–368. Maribor: Pokrajinski arhiv Maribor.