

1.01 Izvirni znanstveni članek

UDK 004.057.3:930.25

Prejeto: 25. 4. 2012

Analiza primernosti formata PDF/A za dolgoročno hrambo gradiva v digitalni obliki

TATJANA HAJTNIK

mag., univ. dipl. ing. računalništva, vodja Sektorja za e-arhive in računalniško podporo
Arhiv Republike Slovenije, Zvezdarska 1, SI-1000 Ljubljana
e-pošta: tatjana.hajtnik@gov.si

IZVLEČEK

Ob vedno večjih količinah gradiva, ki že izvirno nastaja v digitalni obliki, intenzivno prihaja v ospredje vprašanje, kako to gradivo tudi dolgoročno hraniti, da bo ostalo dostopno, uporabno, celovito in sveda verodostojno. Trenutno najbolj razširjena strategija reševanja tega vprašanja je pretvorba gradiva iz ene digitalne oblike v drugo, ta pa mora biti primerna za dolgoročno hrambo. Eden najbolj uveljavljenih formatov v svetu za dolgoročno hrambo besedilnih in mešanih dokumentov je format PDF/A. V Sloveniji je s predpisi na področju upravljanja z dokumenti prav tako na seznamu priporočenih formatov. Preučili smo različne vire in naredili testni primer pretvorbe besedilnega dokumenta iz njegove izvirne oblike v PDF/A. S pomočjo tega smo utemeljili, da format PDF/A ne izpolnjuje vseh kriterijev, ki jih za formate, primerne za dolgoročno hrambo, določajo slovenski predpisi.

KLJUČNE BESEDE: dokumentarno in arhivsko gradivo v digitalni obliki, dolgoročna digitalna hramba, izvirna oblika zapisa, pretvorba, PDF/A, format, oblika zapisa

ABSTRACT

ANALYSIS OF PDF/A AND ITS SUITABILITY FOR LONG-TERM PRESERVATION OF ELECTRONIC RECORDS

Faced with the ever-increasing number of records originally created in digital form, experts around the world are dealing with the issue of how to enable the long-term preservation of such records and at the same time maintain the records' availability, usability, integrity and authenticity. Currently, the most widely used strategy for the long-term preservation of digital documents is their conversion from the original format into one that is more suitable for such preservation. PDF/A is one of the formats most commonly used for the long-term preservation of textual and mixed documents in the world. In Slovenia, it is also on the list of the formats recommended for this purpose by the Slovenian records management legislation. We studied a variety of sources and did a test conversion of a textual document from its original form into PDF/A. Based on that, we were able to establish that PDF/A does not meet all the criteria defined by Slovenian regulations as necessary for a suitable long-term preservation format.

KEY WORDS: documents and archives in digital form, long term digital preservation, original format, conversion, PDF/A, formats

1. Uvod

Na eni strani pri našem vsakdanjem delu nastajajo vedno večje količine gradiva v digitalni obliki, po drugi strani pa nam predstavlja velike izzive hranjenje. Ti so še posebno izraziti takrat, ko te dokumente želimo ali moramo ohraniti za daljši čas, in to na način, da bodo dostopni in uporabni, da bomo zaupali v njihovo verodostojnost in jih po potrebi morda tudi ponovno uporabili. S tem izzivom se predvsem v zadnjih letih intenzivno ukvarjajo tako rekoč po vsem sveru, še posebej izrazito pa iščejo odgovore na te izzive arhivisti in bibliotekarji ter z njimi tudi informatiki. To nam ne nazadnje dokazujejo tudi različna raziskovalna dela, članki, doktorske disertacije in veliko število organiziranih dogodkov, od seminarjev do konferenc. Med najbolj izpostavljenimi tveganji pri hranjenju gradiva v digitalni obliki za daljša časovna obdobja so tista, ki so povezana predvsem z nosilci in oblikami zapisov. Strokovnjaki za zmanjšanje teh tveganj ponujajo več možnih rešitev, v današnjem času pa je največ govora o strategijah pri pretvorbi iz ene digitalne oblike v drugo kot načinu, ki bo zagotavljal dostopnost in uporabnost tega gradiva za ves čas, kolikor ga je treba hraniti. Zato nastajajo tudi različni standardi za oblike zapisa, ki naj bi bili primerni za daljšo hrambo gradiva. Tudi v Sloveniji se je stroka (arhivisti, bibliotekarji, informatiki) že pred nekaj leti spoprijela s temi izzivi in zato celo predlagala spremembo oz. dopolnitve predpisov na področju arhivske stroke.¹ V okviru le-teh je kot obvezno poliniko dolgoročnega hranjenja gradiva v digitalni obliki določila pretvarjanje gradiva iz ene digitalne oblike v drugo. Ob tem je priporočila tudi uporabo standardiziranih oblik zapisa. Te morajo glede na svoje značilnosti ustrezati kriterijem, ki jih je stroka opredelila kot nujne, če jih želimo uporabljati pri hrambi gradiva v digitalni obliki dalj časa. Danes je na voljo veliko standardiziranih oblik zapisa za različne tipe gradiva, ki poskušajo odgovorjati na izzive dolgoročnega hranjenja gradiva. Glede na dostopne vire pa, kot kaže, še ne obstaja takšna, ki bi odgovorila na vsa vprašanja, povezana z dolgotrajnejšim hranjenjem gradiva v digitalni obliki. V prispevku se bomo ukvarjali z vprašanjem, ali je temu res tako.

Za nalogo smo si žadali potrditi ali ovrči te hipoteze:

1. *pretvorba dokumenta iz ene digitalne oblike v drugo lahko pomembno vpliva na njegovo dostopnost, uporabnost, celovitost in avtentičnost;*
2. *standard PDF/A ne izpolnjuje določil slovenskih predpisov za obliko zapisa, ki je primerna za dolgoročno hrambo besedilnih in mešanih dokumentov;*
3. *pretvorba dokumenta iz ene digitalne oblike v drugo ni dovolj dobra strategija za dolgoročno hranjenje.*

Osrednjo pozornost bomo namenili predvsem obliki zapisa, ki je zdaj v sveru med najbolj razširjenimi na področju dolgoročnega hranjenja gradiva v digitalni obliki, to je PDF/A.

Zato smo najprej pregledali in analizirali vire, ki obravnavajo načela zagotavljanja dolgoročne hrambe gradiva v digitalni obliki. Nato nas je zanimalo, katere lastnosti mora izpolnjevati standard, ki ga slovenski predpisi priporočajo kot primerne za dolgoročno hrambo gradiva v digitalni obliki. S pregledom in analizo virov, ki obravnavajo PDF/A kot standard, primeren za dolgoročno hrambo gradiva v digitalni obliki, smo ugotavljali, kako je ta standard razširjen v drugih državah. Z metodo primerjanja (komparativna metoda) smo želeli potrditi ali zavreči hipotezo, da PDF/A ne izpolnjuje zahtev, ki jih slovenska zakonodaja določa za standarde, primerne za dolgoročno hrambo besedilnih in mešanih dokumentov v digitalni obliki. V sklepu pa smo tudi ugotavljali, ali je pretvorba dokumenta iz ene digitalne oblike v drugo dovolj dobra strategija za dolgoročno hranjenje.

V okviru raziskave obravnavamo digitalno gradivo, ki je izvorno nastalo v elektronskem okolju in ga je treba tam tudi dolgoročno hraniti (izvorno digitalno gradivo). Omejili smo se samo na besedilne in mešane dokumente.

2. Gradivo v digitalni obliki in problem dolgoročnega hranjenja

Vse od uvedbe računalnikov v petdesetih letih prejšnjega stoletja smo priča hitremu razvoju zmogljivosti računalniške tehnologije (strojne in programske opreme, komunikacij, nosilcev podatkov itd.).² Zaradi informacijske tehnologije, danes vedno bolj zmogljive in vse preprostejše za uporabo, omogočajo ustvarjanje vedno večjih količin gradiva v digitalni obliki. To gradivo pa je zaradi različnih razlogov treba, in ga bo treba tudi v prihodnje, ohraniti. Namen te hrambe je zagotoviti

¹ Do leta 2006 je veljal Zakon o arhivskem gradivu in arhivih /ZAGA/ (Ur. l. RS, št. 20/97, 32/97 in 24/2003), z marcem 2006 pa je začel veljati Zakon o varstvu dokumentarnega in arhivskega gradiva ter arhivih /ZVIDAGA/ (Ur. l. RS, št. 30/2006), ki ureja tudi varstvo gradiva v digitalni obliki.

² ICA, *Guide for managing electronic records from an Archival Perspective*, str. 15.

trajen dostop do gradiva za sedanje in prihodnje generacije.³ Duranti in Thomassen⁴ pravita, da se čas, za katerega je treba gradivo ohraniti, lahko razlikuje od meseca do sto let. V virih najdemo različne definicije hrambe dokumentarnega gradiva, pogosto je uporabljen tudi pojem »digitalna hramba«. Tako je npr. v Wikipediji, ki povzema definicije posameznih strokovnjakov⁵ za hranjenje gradiva v digitalni obliki in različnih projektov, digitalna hramba (angl. *Digital preservation*) definirana kot »vrsta postopkov, dejavnosti in upravljanje digitalnih informacij v daljšem časovnem obdobju, da je zagotovljena dolgoročna dostopnost le-teh«. Cilj digitalne hrambe je hranjenje gradiva, ki je nastalo s pretvorbo, in še zlasti tistega gradiva, ki je nastalo v digitalni obliki brez analognega dvojnika. Zaradi razmeroma kratke življenjske dobe digitalnih informacij je namreč ohranjanje stalen proces.

Vprašanje dolgoročnega hranjenja digitalnega gradiva je navzoče že vrsto let. Kot odgovorni za arhivsko gradivo so se predvsem arhivisti in bibliotekarji že zelo zgodaj začeli zavedati težav pri hranjenju digitalnih informacij za prihodnje generacije.⁶ Najbolj znan referenčni model »Odprti arhivski informacijski sistem«, ki je hkrati tudi standard ISO 14721:2003, dolgoročno hrambo definira kot »dolgoročno shranjevanje digitalne informacije brez napak z možnostjo ponovne pridobitve (angl. *retrieval*) in interpretacije (angl. *interpretation*) za ves čas hrambe te informacije».⁷ Izraz »long-term«⁸ je definiran kot »dovolj dolgo ukvarjanje z vplivi spreminjanja tehnologij vključno s podporo za nove medije in formate podatkov ali s spremembami v skupnosti uporabnikov. Dolgoročno se lahko podaljša za nedoločen čas«. Pri tem »ponovna pridobitev«⁹ pomeni pridobitev potrebnih, dolgoročno hranjenih digitalnih dokumentov brez napak in morebitnih poškodb. »Interpretacija«¹⁰ pa pomeni, da bodo ponovno pridobljeni dokumenti npr. besedilni, slikovni ali zvočni dokumenti, preoblikovani in prikazani v uporabni obliki. To pogosto razlagamo kot »vračanje«¹¹ na način, da bodo ljudem dostopni. Vendar pa bo v številnih pri-

merih to pomenilo, da se dokumenti lahko obdelujejo le s pomočjo računalnikov.⁹

Da je digitalna hramba osnovna težava informacijske družbe, saj količina informacij eksponentno raste, vsebina pa je vedno bolj raznovrstna, je leta 2005 opozorila tudi komisija evropskih skupnosti (*evropska komisija*).¹⁰ Že raket je ugotavljala, da je izkušenj z digitalno hrambo malo, da se na eni strani zakonski okvir razvijajo, vendar pa so viri nezadostni in izid prizadevanj za digitalno hrambo negotov.¹¹ V sporočilu evropskemu parlamentu, svetu, evropskemu ekonomsko-socialnemu odboru in odboru regij (v nadaljevanju *Sporočilo*) je zapisala, da obstajajo različni vzroki za izgubo digitalnih vsebin.

Kot prvi vzrok omenja menjavanje generacij strojne opreme, zato lahko dokumenti postanejo neberljivi.

Kot primer evropska komisija v Sporočilu navaja: »Ob 900-letnici knjige Domesday book leta 1985 je bila sestavljena nova, multimedijška izdaja. Leta 2002 se je zdelo, da je disk postal neberljiv, saj so postali računalniki, sposobni brati tak format, redki. Da bi ga rešili, so razvili sistem, ki je lahko do diskov dostopal z emulacijsko tehniko. Zanimivo je, da se pojavljajo težave z dostopanjem do digitalnih podatkov iz leta 1986, original knjige Domesday Book pa je po več kot 900 letih še vedno uporaben.«

Kot drugi vzrok evropska komisija v *Sporočilu* navaja hitro menjavanje in zastarelost programske opreme, s katero je gradivo ustvarjeno, s katero ga ra upravlja, in hranjeno. Če podatki niso preneseni v zdaj veljavne programe oziroma ni poskrbljeno za hrambo izvornega besedila, lahko postane iznos podatkov zelo drag ali celo nemogoč. To posebej velja za »zaprt«¹² formate podatkov, za katere izvorno besedilo navadno ni splošno poznano. Za dodatni razlog izgube digitalne vsebine navajajo tudi omejeno življenjsko dobo nosilcev za digitalno

³ Conway: *Archival Preservation Practice in a Nationwide Context*, str. 206.

⁴ Duranti: *Concepts, Principles, and Methods for the Management of Electronic Records*, str. 271–279; Thomassen: *A First Introduction to Archival Science*, str. 373–385.

⁵ Prim. Consultative Committee for Space Data Systems (CCSDS), Margaret Hedstrom, Jeff Rothenberg.

⁶ Hedstrom: *Understanding Electronic Records: A Framework for Research on Electronic Records*, str. 334–354.

⁷ Prim. ISO 14721:2003. *Recommendation for space data system standards*, stran 1.

⁸ Prav tam, stran 1–8.

⁹ Prim. *Digital preservation*.

¹⁰ Komisija Evropskih skupnosti (evropska komisija) je izvršilni organ Evropske unije. Sestavlja jo 27 resorjev vključno s predsednikom. Vsaka država ima v komisiji po enega komisarja.

¹¹ Prim. *Sporočilo Komisije Evropskemu Parlamentu, Svetu, Evropskemu Ekonomsko-Socialnemu Odboru in Odboru Regij - iz 2010: digitalne knjižnice*.

¹² Zaprti (lastniški) formati: uvede in kontrolira jih eno podjetje in so lahko zaščiteni s patentom, lahko se samovoljno spreminjajo, lahko so zelo pomanjkljivo dokumentirani. Lahko pa so kljub vsemu tudi splošno razširjeni zaradi tržne moči njihovega lastnika, niso pa tudi uradno odobreni – če jih ne odobri neodvisni organ za standarde, npr. ISO. Glej: *Proprietary standard*.

hrambo (npr. CD-ROM-ov). Kot osnovne izzive digitalne hrambe Evropska komisija izpostavlja predvsem finančne, organizacijske, pravne in tehnične izzive. Slednje vidi predvsem v načinu razumevanja dolgoročnega hranjenja velike količine hitro se spreminjajočih podatkov v obrotu.¹³

3. Dolgoročna hramba gradiva v digitalni obliki na podlagi slovenskih predpisov

Zaradi potrebe po urejeni in v skladu s tehničnimi standardi učinkoviti hrambi dokumentarnega in arhivskega gradiva v digitalni obliki (pa naj gre za gradivo, ki je nastalo v izvorno digitalni obliki oziroma je bilo digitalizirano iz drugih oblik) ter dokazne vrednosti hranjenih dokumentov v digitalni obliki pred državnimi in sodnimi organi, je državni zbor Republike Slovenije lera 2006 na podlagi dopolnitev obstoječega arhivskega zakona¹⁴ sprejel novega,¹⁵ s katerim ureja tudi varstvo dokumentarnega in arhivskega gradiva v digitalni obliki (v nadaljnjem besedilu: *gradivo v digitalni obliki*). Zakon ureja način, organizacijo, infrastrukturo in izvedbo hrambe dokumentarnega in arhivskega gradiva v fizični in digitalni obliki vključno s pravnimi učinki takšne hrambe (sistem pravne veljavnosti gradiva v digitalni obliki). V procesu hrambe, še posebej dolgoročne, imajo pomembno vlogo predvsem informacijsko tehnološka sredstva, v prvi vrsti programska in strojna oprema ter nosilci in oblike zapisa. Že zakon¹⁶ določa splošne pogoje, ki jih mora izpolnjevati strojna in programska oprema, njemu pripadajoča uredba¹⁷ pa natančneje opredeljuje tudi zahteve za oblike¹⁸ in nosilce zapisa¹⁹ ter ukrepe za zmanjšanje tveganj, povezanih z zagotavljanjem dostopnosti, celovitosti, avtentičnosti in uporabnosti gradiva oz. njegove vsebine.²⁰ Natančneje poslovne, organizacijske in tehnološke razmere za zagotavljanje dolgoročne hrambe, ki bo rudi skladna z zakonom, je državni arhiv (Arhiv Republike Slovenije, v nadaljevanju: *Arhiv RS*) na podlagi zakona²¹ objavil v dokumentu *Enotne tehnološke zahteve*.²² Bistvo je predvsem v poenotenju praks in

tehnoloških postopkov za upravljanje ter dolgoročno hrambo gradiva v digitalni obliki.

Z VSDAGA je »dokumentarno gradivo« definirano kot »izvirno in reproducirano (pisano, risano, tiskano, fotografirano, filmeno, fonografirano, magnetno, optično ali kako drugače zapisano) gradivo«. *Arhivsko gradivo* pa je del dokumentarnega gradiva, ki ima »trajen pomen za znanost in kulturo ali trajen pomen za pravno varnost oseb v skladu s strokovnimi navodili pristojnih arhivov«.²³ V nadaljevanju je Arhiv RS z ETZ »gradivo v digitalni obliki« definiral kot »izvirno digitalno gradivo« (izvirno je nastalo v elektronskem okolju in ga je treba tam dolgoročno hraniti) in »digitalizirano gradivo« (izvirno je nastalo v fizični ali analogni obliki, s postopki digitalizacije pa sta vsebina in oblika dokumentov pretvorjeni v digitalno obliko).²⁴

VSDAGA hrambo dokumentarnega gradiva definira kot »hranjenje izvirnega dokumentarnega gradiva ali uporabnosti vsebine tega gradiva«.²⁵ Na podlagi te opredelitve je hrambi izvirnega dokumentarnega gradiva enaka hramba zajetega gradiva, če zagotavlja zajetemu gradivu vse učinke izvirnega gradiva (uporabnost vsebine gradiva). Zakon »dolgoročno hrambo« obravnava kot hrambo gradiva več kot pet let.²⁶

Cilj vsake hrambe naj bi bil zagotavljanje izpolnjevanja potreb, zaradi katerih je gradivo nastalo ali ga hranimo.²⁷ Mednje je mogoče šteti najmanj:

- dokazovanje pravnih, finančnih in podobnih stanj,
- dokazovanje sprejemanja in izvajanja poslovnih odločitev,
- dokazovanje izvedbe in načina izvajanja poslovnih procesov in
- ohranitev kulturne dediščine v obliki vseh vrst zapisov.

V prvi vrsti naj bi bili namreč učinki gradiva povezani s pravnim varstvom oseb, o tem pa priča tudi historično dejstvo, da so najstarejša gradiva nastala zaradi pravne varnosti tistih, ki so hranili dokumente. Te učinke VSDAGA opredeljuje z načeli ohranjanja uporabnosti vsebine gradiva in njegove celovitosti, posledica tega pa je tudi ohranjanje avtentičnosti oziroma verodostojnosti. Pri hranjenju gradiva v digitalni obliki na način, da bodo ves čas hrambe zagotovljeni učinki, zaradi ka-

¹³ *Sporočilo Komisije Evropskemu Parlamentu, Svetu, Evropskemu Ekonomsko-Socialnemu Odбору in Odborni Regiji - iz 2010: digitalne knjižnice*, točka 6.

¹⁴ ZAGA, Ur. l. RS, št. 20/1997.

¹⁵ VSDAGA, Ur. l. RS, št. 30/2006.

¹⁶ Prav tam, 71. člen.

¹⁷ VSDAGA, Ur. l. RS, št. 86/2006.

¹⁸ Prav tam, 14. člen.

¹⁹ Prav tam, 15. člen.

²⁰ Prav tam, 16.–18. člen.

²¹ VSDAGA, 91. člen.

²² *Enotne tehnološke zahteve ver 2.0*, I–III.

²³ VSDAGA, 2. člen.

²⁴ ETZ 2.0 - 1. del: *Uvodna poglavja*, stran 9.

²⁵ VSDAGA, 3. člen.

²⁶ Prav tam, 2. člen.

²⁷ Prim. *Arhiviranje, hranjenje in upravljanje dokumentov*, poglavje 3.

terih je gradivo dejansko nastalo, govorimo o »**učinkoviti²⁸ hrambi gradiva**«. Pri pomenu te besedne zveze se ni mogoče neposredno opirati na razlago ali definicijo le-te v ZVDAGA ali podrejenih dokumentih, saj jo je mogoče najti le med določili, ki opredeljujejo namen, cilj oziroma učinek posameznega poslovnega procesa.²⁹ Merwe in Deventer³⁰ navajata, da bo strategija digitalne hrambe učinkovita, ko ne bo le prepoznala tveganja temveč bo predstavila tudi izvedljive in cenovno dostopne rešitve. Po Hedstromovi in Montgomeryju³¹ pa učinkovita digitalna hramba zahteva upravljanje življenjskega cikla digitalnih podatkov vse od njihovega nastanka do hrambe, pretvorbe in zagotavljanja rednega dostopa.

Učinkovita hramba gradiva: *z hrambo celotnega besedila ZVDAGA in podrejenih predpisov jo je mogoče pojasniti predvsem na podlagi temeljnih načel zakona: trajnosti, dostopnosti, uporabnosti, celovitosti in avtentičnosti (verodostojnosti) gradiva v digitalni obliki.*

Med temeljna načela dolgoročne hrambe sodi po ZVDAGA tudi načelo varstva kulturnega spomenika,³² ki pa ga tesno povezujemo s hrambo gradiva s statusom arhivskega gradiva. Tak status ima zaradi vsebine, oblike ali okoliščin, v katerih je nastalo oziroma je bilo v uporabi, in ima lastnosti kulturne dediščine.

Za vse gradivo, ki je predmet učinkovite hrambe, morajo biti ta načela zagotovljena. To velja ne le za gradivo, ki ga je treba hraniti trajno, ampak seveda tudi za tiste oblike gradiva, ki imajo krajše roke hrambe.

V naši raziskavi obravnavamo gradivo, ki je v izvorni obliki digitalno in ga je treba hraniti dolgoročno. Sem sodi gradivo, ki je glede na rok hrambe trajno³³ ali arhivsko gradivo³⁴ ter preostalo

gradivo, ki ga je treba hraniti več kot per let. Ne obravnavamo pa gradiva, ki izvorno nastane v fizični³⁵ ali analogni obliki,³⁶ njegova vsebina in oblika pa sta v postopku digitalizacije³⁷ pretvorjeni v digitalno obliko.

3.1 Dokument v digitalni obliki in načela dolgoročne hrambe

Kot temeljno enoto gradiva v digitalni obliki bomo obravnavali dokument, za katerega pa v akademskih in strokovnih krogih obstaja več definicij.³⁸ Mednarodni standard za upravljanje dokumentov v elektronski obliki ISO 15489-1 dokument definira kot »*zapisano informacijo ali predmet, ki ga je mogoče obravnavati kot enoto*«³⁹ Mednarodni arhivski svet pa v Smernicah za upravljanje elektronskih dokumentov iz arhivske perspektive (v nadaljevanju *Smernice*) dokument definira kot »*zapisano informacijo, ki je ustvarjena ali prejeta; je odraz postopka ali zaključek aktivnosti posameznika ali ustanove ter zajema zadostno vsebino, kontekst in strukturo za zagotovitev dokaza o dejanju*«⁴⁰

²⁸ Dokumentarno gradivo v fizični obliki je dokumentarno gradivo na fizičnem nosilcu zapisa, ki omogoča reprodukcijo vsebine brez uporabe informacijsko komunikacijskih ali sorodnih tehnologij (npr. na papirju, filmu, itd.) (ZVDAGA, 2. člen).

²⁹ Dokumentarno gradivo v analogni obliki (npr. analogni audio/video zapis) je dokumentarno gradivo v analogni obliki zapisa in shranjeno na elektronskem nosilcu zapisa. (ZVDAGA, 2. člen).

³⁰ Digitalizacija je strokovni tehnološko-organizacijski postopek pretvarjanja vsebin in pojavne oblike dokumentov iz njihove fizične v digitalno obliko. Prim. ITZ 2.0, 1. del: Uradna paglaja, str. 9.

³¹ Primeri definicij dokumenta:

»Dokument je zapisana informacija, ustvarjena ali sprejeta v okviru institucionalnih ali individualnih dejavnosti in vključuje vsebino, kontekst in strukturo v zadostnem obsegu za zavarovanje dokazov o tej dejavnosti« (Prim. ICA, *Guide for managing electronic records from an Archival Perspective*, str. 22).

»Dokumenti so zapisi, ki so zanesljivi in popolni, to je, so sposobni podajati informacije, se lahko uporabljajo v transakcijah in so uporabljeni za doseganje namena, zaradi katerega so bili ustvarjeni« (Prim. Duranti: *Diplomatics: New Uses for an Old Science*, str. 74).

»Dokumenti so, ko so združeni z ustreznimi metapodatki, dokaz o njihovi vsebini, obliki in kontekstu« (Prim. Bearman: *Capturing Records' Metadata*, str. 272).

³² ISO 15489-1, *Information and Document Management – Records Management (Part 1: General)*, str. 3.

³³ Prim. ICA, *Guide for managing electronic records from an Archival Perspective*.

³⁴ Pojem »učinkovita« je v Slovenskem slovarju knjižnega jezika razložen: »s katerim se dosega želeni, pričakovani cilj«.

²⁹ Prim. ZVDAGA, 8. alineja 2. člena in 8. člen; UVDAGA, 2. odstavek 9. člena.

³⁰ Merwe & Deventer: *Planning an effective digital preservation strategy*, str. 6.

³¹ Hedstrom & Montgomery: *Digital Preservation Needs and Requirements*, str. 20.

³² Prav tam, 7. člen.

³³ *Trajno gradivo* ima lahko trajen pomen za organizacijo, pri kateri je nastalo, ali pa ga je treba trajno hraniti, ker tako določajo zakonski predpisi.

³⁴ *Arhivsko gradivo* je po definiciji slovenskih predpisov tisti del dokumentarnega gradiva (predvsem javnopravnih oseb, nekaj pa je tudi zasebnega arhivskega gradiva), ki ima trajen pomen za znanost in kulturo ali pravno varnost države, njenih institucij in posameznikov ter ga je kot takega treba hraniti trajno. (ZVDAGA, 2. člen).

Dejansko gre za precej široko definicijo, ki lahko zajema različne vrste dokumentov. Ti se namreč lahko pojavljajo v različnih oblikah in predstavitev. Navadno so predstavljeni kot logično razmejeni informacijski objekti, na primer različni zapisi. Vendar pa vedno pogosteje najdemo zapise v obliki distribuiranih objektov, kot so relacijske baze podatkov in sestavljene zapise. *Smernice* dokumente razvrščajo na podlagi dveh meril:⁴¹

- *njihove funkcije*: to je povezave datotek z različnimi vrstami aktivnosti in transakcij v pisarniškem okolju. Primeri vključujejo zadeve, sodne spise, dosjeje (glede na določeno dejavnost), personalne zadeve, korespondenco, spletne dokumente, itd.; oziroma
- *njihove oblike in formata*: primeri vključujejo zapise, oblikovane v urejevalniku besedil (npr. MS Word), podatkovne zbirke, zapise v hipertekstu, slike, preglednice, elektronska sporočila, zvočna sporočila, video, itd.

Smernice poudarjajo, da morajo organizacije zaje-
ti in vzdrževati dokumente, ki nastajajo pri izvaja-
nju njihovih poslovnih funkcij za zadovoljitev po-
slovnih potreb in zakonskih zahtev. Iz te per-
spektive je »glavni namen oblikovanja dokumentov
in njihovega hranjenja zagotoviti dokaz« za delo-
vanje organizacije ali odgovornost pravne osebe ali
posameznika. Za podporo poslovnim funkcijam in
zagotovitev dokaza mora imeti dokument določene
lastnosti in *Smernice* poudarjajo dve:

- *avtentičnost (verodostojnost)*: opredeljena je kot »hra-
njenje izvirnih lastnosti dokumenta v daljšem
časovnem obdobju glede na kontekst, strukturo
in vsebino«, to pomeni, da je dokument tisto,
kar naj bi bil;
- *zanesljivost*: opredeljena je kot sposobnost doku-
menta, da »služi kot zanesljiv dokaz«, nanašajoč
se na izvor in verodostojnost dokumentov kot
dokaz.

Tudi drugi viri, vključno z Mednarodnim stan-
dardom za upravljanje dokumentov ISO 15489-1,
so potrdili pomen teh kvalitete in dodali še dve
tesno povezani lastnosti dokumenta:

- *celovitost*: nanaša se na dejstvo, da je dokument
celovit in nespremenjen;
- *uporabnost*: opredeljena je kot sposobnost za to,
da je mogoče najti, priklicati, predstaviti in in-
terpretirati dokument.⁴²

Namreč le dokumenti z navedenimi značil-
nostmi bodo imeli dovolj vsebine, strukture in kon-
teksta za zagotovitev celovitega pregleda aktivnosti
in transakcij, na katere se nanašajo ter odražajo od-
ločitve, aktivnosti in odgovornosti. Če so taki do-
kumenti vzdrževani na dostopen, razumljiv in upo-
raben način, bodo lahko po preteku določenega ča-
sa podpirali poslovne potrebe in bodo uporabljeni
v zvezi z odgovornostjo. Na podlagi teh lastnosti je
Mednarodni arhivski svet opredelil osnovne zahtev-
te za dolgoročno hrambo dokumentov, in sicer
morajo le-ti biti:⁴³

- dostopni in razumljivi,
- celoviti,
- avtentični,
- primerni za obdelavo in
- potencialno ponovno uporabni.

V nadaljevanju na kratko povzemamo opredelit-
ve navedenih zahtev:

Dokument je **dostopen**, če imamo tehnologijo,
ki omogoča, da ga lociramo (dostop do dokumen-
ta) in potem prevedemo v obliko, ki jo človeški čuti
lahko zaznavajo – dostop do vsebine dokumenta.
Pri tem seveda ni nujno, da ima dokument vse
lastnosti programske opreme, ki ga je ustvarila.

Razumljivost dokumenta bomo obravnavali v
smislu, ali lahko ugotovimo pomen dokumenta in
pomen, ki ga dokument želi prenesti. Dokument je
sam po sebi že lahko razumljiv, če je dostopen. V
to kategorijo bo sicer spadalo veliko dokumentov
ob predpostavki, da razumemo jezik, v katerem so
bili dokumenti napisani. Kljub temu pa predpo-
stavljamo, da bodo že na ravni razumevanja doku-
menta težave. Pogosto je namreč poleg samih do-
kumentov treba imeti tudi dodatne podatke, npr.
šifrant, makroje ali formule. Če ti ne bodo ohran-
jeni in povezani z osnovnimi dokumenti, bo ra-
zumevanje osnovnih dokumentov oteženo.

Celovitost bomo obravnavali kot lastnost na-
bora dokumentov in ne samo kot postavko znotraj
dokumenta, čeprav jo je mogoče aplicirati na oboje.
Da bo nabor dokumentov celovit, moramo biti
sposobni dokazati, da mu nobena postavka ni bila
odvzeta drugače kot skladno s pravili, posravljenimi
za ta nabor.

Za vzpostavitev **avtentičnosti** dokumenta mo-
ramo biti sposobni dokazati, da je ohranjeno to, kar
smo vedeli o njem, ko smo ga prejeli oz. ustvarili;
da je bil dokument prejet ali ustvarjen takrat, ko

⁴¹ Prav tam, stran 10.

⁴² Prim. ISO 15489-1, *Information and Document Management – Records Management (Part 1: General)*, poglavje 3.12.

⁴³ Prim. ICA, *Guide for managing electronic records from an Archival Perspective*, str. 34–35.

trdimo, da je bil, da je bil proces, v katerem je bil oblikovan (ročno ali avtomatizirano), prav tisti, za katerega trdimo, da je njegov del; da vsebina dokumenta ni bila na noben način spremenjena, odkar je dokument postal del sistema za hrambo.

Da dokumente lahko imamo kot **primerne za obdelavo**, moramo biti sposobni ravnati z njimi, jih izbrati in prikazati z uporabo kriterijev, primeranih za namen hrambe. To lahko pomeni uporabo opreme, podobne ali enake tisti, ki je obstajala v izvornem sistemu za oblikovanje dokumentov. Vendar ima v številnih primerih sistem, v katerem so bili dokumenti oblikovani, lahko funkcije, ki jih pri dolgoročni hrambi – dostopnosti ne bomo potrebovali. Npr. zbirka ohranjenih dokumentov, narejenih z urejevalnikom besedil, mora samo zagotoviti način za lociranje pomembnih dokumentov ter za ogled njihove vsebine na zaslonu, natisnjenih kopij ali na drug primeren način. Ni treba ohraniti sposobnosti za urejanje ali izvedbo drugih akcij, ki bi jih omogočal izvorni urejevalnik besedila.

Dokumenti so **potencialno ponovno uporabni**, če je iz njih mogoče izluščiti informacije ali kako drugače omogočiti dokumentu, da sodeluje s sodobnimi informacijskimi sistemi. To je strožja zahteva kot zahteva po dostopnosti in možnosti obdelave. Zahtevi (primernost za obdelavo in morebitno ponovno uporabo) je sicer mogoče izpolniti z ohranitvijo izvorne strojne in programske opreme, na kateri so bili dokumenti oblikovani. Zahtevo po ponovni uporabi je sicer mogoče tudi zanemariti, saj je v večini primerov ne bomo potrebovali, v določenih primerih pa jo bo vendarle zaželeno doseči. Na tak način sta navadno zagotovljeni tudi dostopnost in razumljivost dokumentov.

ZVDA GA je kot temeljna načela dolgoročnega branjenja dokumentov v digitalni obliki, opredelil:

- **uporabnost:** *zmožnost za reprodukcijo in primernost reprodukcije za uporabo ves čas trajanja hrambe (npr. branje, tiskanje dokumenta);*
- **trajnost:** *hramba dokumentarnega gradiva mora zagotavljati trajnost tega gradiva oziroma trajnost reprodukcije njegove vsebine;*
- **celovitost:** *zagotovljene morajo biti pravilnost, nespremenljivost in popolnost gradiva oziroma reprodukcije njegove vsebine, urejenost gradiva oziroma njegove vsebine ter dokazljivost izvora gradiva (provenience) ves čas hrambe;*
- **dostopnost:** *gradivo oziroma reprodukcija njegove vsebine mora biti ves čas trajanja hrambe zavarovana pred izgubo ali okrnitvijo celovitosti ter dostopna tistim, ki imajo za to ustrezna pooblastila.*

Šele v nadaljevanju zakon pri opredelitvi varne hrambe zajetega gradiva v digitalni obliki poleg zahteve po uporabnosti, trajnosti, celovitosti in dostopnosti izpostavi tudi zahtevo po **avtentičnosti**. T.e.to tolmači kot dokazljivo povezanost reproducirane vsebine z vsebino izvirnega gradiva oziroma izvorom tega gradiva.⁴⁴ Če bodo pri hrambi izvirnega gradiva v digitalni obliki zagotovljene dostopnost, uporabnost in celovitost, ZVDA GA takšno hrambo opredeljuje kot *»varno hrambo«*.⁴⁵

4. Programska oprema in oblika zapisa

4.1 Programska oprema

Programska oprema (angl. software,⁴⁶ v slovenščini tudi softver ali programje) je skupek računalniških programov, ki skupaj s strojno opremo računalnika sestavlja celoto. Z njo ustvarjamo tudi različne vrste dokumentov. Glede na namen in funkcijo pa obstajajo tudi različne vrste programske opreme:

- *sistemska programska oprema:* skrbi za usklajeno delovanje posameznih komponent strojne opreme, komunikacijo med uporabnikom in računalnikom, komunikacijo med računalnikom in drugimi računalniškimi napravami ter izvajanje aplikacijske programske opreme. Med sistemsko programsko opremo sodijo npr. operacijski sistemi, pogonski programi, razvojna orodja oz. programski prevajalniki, servisni ali uslužnostni programi;
- *aplikacijska programska oprema:* z njeno pomočjo uporabniki izvajajo svoje specifične naloge in v okviru tega ustvarjajo različne vrste dokumentov. K tej programski opremi sodijo npr.:
- *urejevalniki besedil:* uporabniku omogočajo pisanje in oblikovanje različnih vrst dokumentov oz. besedil (npr. Wordstar, Word Perfect, Microsoft Word, WordPad, Notepad);
- *preglednice:* temeljijo na računovodskih načelih in uporabi matematičnih, statističnih in logičnih operacij; z njimi uporabnik izdeluje tabele, diagrame, izdeluje finančne izračune (npr. Lotus 1-2-3, Quattro Pro, Microsoft Excel);
- *programi za risanje in obdelavo slik:* omogočajo izdelavo različnih tipov grafičnih prikazov (npr. Corel Draw, Adobe Photoshop, Paint, Paint Shop Pro);

⁴⁴ Prav tam, 27. člen.

⁴⁵ Prav tam, 26. člen.

⁴⁶ Izraz »software« je prvi uporabil John W. Tukey leta 1958. Prim. Leonhardt: John Tukey, 85, Statistician; Coined the Word 'Software'.

- *programi za namizno založništvo (Desktop publishing):* za izvajanje celostnih založniških nalog, ker omogočajo kombiniranje različnih vrst podatkov v celoto, kot je izdelava knjig, brošur, časopisov in revij (Quark Xpress, Adobe InDesign, Microsoft Publisher);
- *programi za delo s podatkovnimi zbirkami (Database):* uporabniku omogočajo hranjenje velikega števila podatkov, tako da je uporaba kasneje kar najhitrejša in najpreprostejša (npr. Paradox, Microsoft Access, Oracle, FoxPro);
- *programi za predstavitev podatkov:* omogočajo izdelavo predstavitev, znotraj katerih je mogoče uporabljati in povezovati podatke različnih vrst (npr. Microsoft PowerPoint, Flash);
- *programi za tehnično risanje, modeliranje in projektiranje (Computer Aided Design):* z njimi uporabnik oblikuje objekte, prikazuje 3D, izdeluje načrte (npr. AutoCAD, Catia);
- *programi za izdelavo spletnih strani* (npr. Adobe GoLive, Dreamweaver, Microsoft Front Page);
- *programi za pošiljanje in prejemanje elektronske pošte* (npr. Microsoft Outlook, Outlook Express, IBM Lotus Notes e-mail);
- *programi za razporedilo, igranje.*

Dokumenti, ustvarjeni s to programsko opremo, ostanejo daljše časovno obdobje enaki, vsaka nova različica programske opreme pa jih interpretira na nekoliko drugačen način. S hitrim tehnološkim razvojem se pojavljajo vedno nove oblike zapisa (primer slika 1) ali pa se nadgrajujejo stare (primer slika 2). Obstaja velika možnost, da s programsko opremo v prihodnosti ne bomo sposobni zagotavljati dostopnosti in uporabnosti dokumentov, zapisanih v starejših oblikah zapisa. Po raziskavi Mind the Gap je prav zastaranje formatov eden največjih izzivov dolgoročne digitalne hrambe.⁴⁷

Digitalno zastaranje torej je in bo tudi v prihodnosti vedno večji problem.

4.2 Izvorna digitalna oblika zapisa (format)

Že hiter pregled različic programske opreme za urejanje besedil v zadnjih tridesetih letih pokaže, za kako hiter tehnološki razvoj in s tem povezano zastaranje oblik zapisa gre. Čeprav so številne različice urejevalnikov besedil nekaj časa veljale za »standarde«, so jih na koncu nadomestile nove. To si lahko pogledamo ob primeru urejevalnika besedil

WordPerfect, ki ga je leta 1982 podjetje WordPerfect Corporation objavilo v različici 1.0a za DOS⁴⁸ operacijski sistem.⁴⁹ Največji tržni delež je dosegel v začetku devetdesetih let prejšnjega stoletja. Razvoj novega operacijskega sistema MS Windows je zahteval ustrezno prilagoditev. Zato je bila na vrhuncu razširjenosti različice DOS urejevalnika besedil Word Perfect v začetku devetdesetih let objavljena tudi nova različica WordPerfect za operacijski sistem Windows 5.1. Kasneje so objavili še različice za različne operacijske sisteme, npr. Linux, Unix, Macintosh, Apple II, Atari ST, Amiga in VAX/VMS.

Ne glede na to, da so se podjetja prilagajala tehnološkemu razvoju operacijskih sistemov in temu prilagajala tudi urejevalnike besedil, pa je urejevalnik besedil Microsoft Word za Windows po njegovi objavi počasi začel izpodrivati preostale, do takrat sicer zelo razširjene urejevalnike besedil (glej Sliko 1).

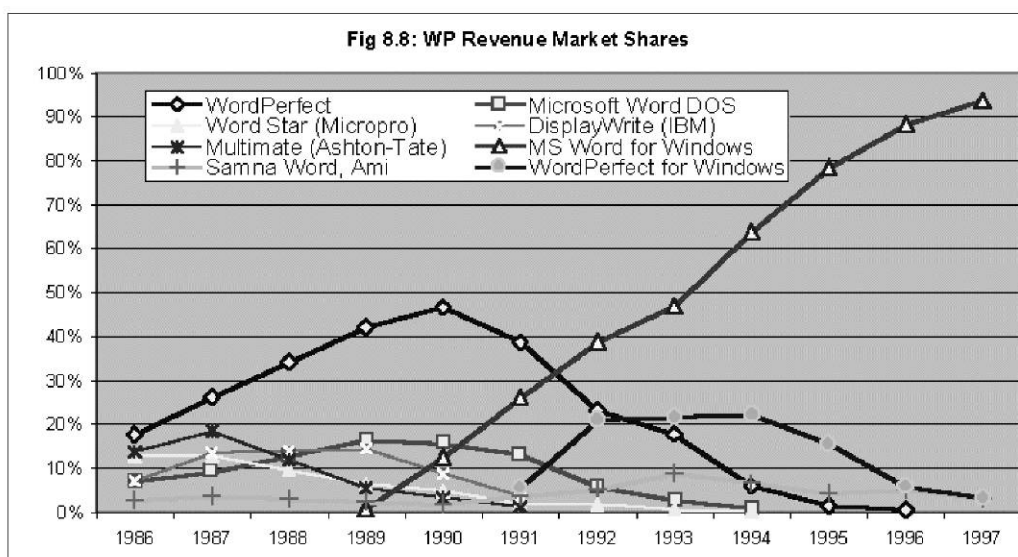
Nove različice kateregakoli urejevalnika besedil predstavljajo določene izboljšave in dodatne funkcionalnosti. Navadno je bilo z njimi mogoče odpirati in urejati tudi dokumente, ustvarjene s starejšimi različicami. Vendar pa poznamo tudi primere, da dokumentov, ki smo jih lahko brali ali urejali z določeno različico urejevalnika besedil, zaradi razvoja systemske programske opreme oz. strojne opreme ni bilo mogoče brati z novejšo različico, čeprav jo je razvilo isto podjetje. Primer so dokumenti, ustvarjeni s starejšo različico Microsoft Works, ki je bila prej Works 4.5. Teh namreč ni več mogoče zagnati v operacijskem sistemu Windows 2000 ali s kasnejšimi verzijami. Takšnim in podobnim primerom se želimo oz. moramo v prihodnosti v kar največji meri izogniti.

Da bomo dokument, ki smo ga ustvarili z določenim urejevalnikom besedil (v tem primeru govorimo o izvorni digitalni obliki zapisa) lahko kasneje tudi odprli in uporabili (brali, tiskali, spreminjali) s programsko opremo, ki jo bomo imeli na razpolago, ga moramo zapisati v posebni obliki zapisa ali formatu.

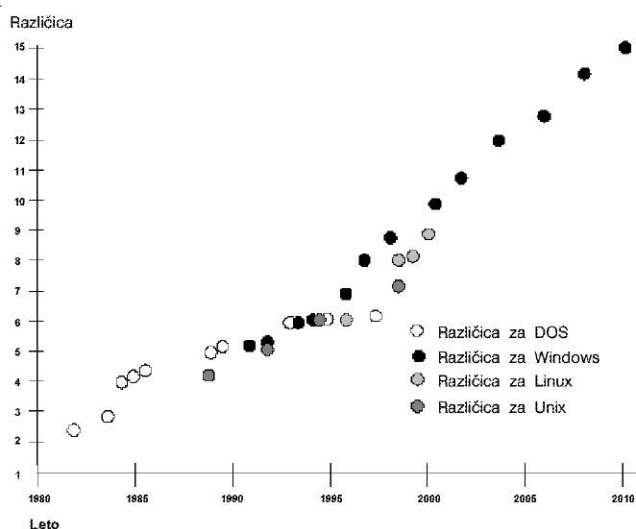
⁴⁸ **DOS (Disk operating system)** je operacijski sistem, razvit za IBM PC in kompatibilne mikroračunalnike. Pri tem je operacijski sistem programska oprema, nujna za delovanje računalnika in deluje kot vmesnik med uporabnikom in strojno opremo računalnika. Leta 1981 je bil izdan na trg **MS-DOS**, ki ga je izdelal Microsoft. V 1980-ih je bil to dominanten operacijski sistem za PC (personal computers) in PC-kompatibilne računalnike.

⁴⁹ Prim. *WordPerfect*.

⁴⁷ Prim. *Mind the Gap: Digital Preservation Needs in the UK*.



Slika 1: Tržni delež urejevalnikov besedil v obdobju 1986–1997
(<http://www.utd.edu/~liebowit/book/wordprocessor/word.html> (16.10.2011))



Slika 2: Zgodovinski razvoj urejevalnika besedil WordPerfect v obdobju 1980–2010
(Izdelano na podlagi podatkov, najdenih 2.11.2011 na naslovu http://en.wikipedia.org/wiki/WordPerfect#cite_note-wpd-0).

Oblika zapisa ali format pomeni reverzibilno kodiran način bitnega zapisa enic in ničel, ki so zapisane na fizičnem nosilcu z ustrezno strojno opremo, ki lahko te bite tudi bere in podaja določeno procesualno in interpretacijsko vrednost za določeno programsko opremo. Ta jo interpretira tako, da je uporabna in razumljiva končnemu uporabniku (človeku, računalniku ipd.). Projekt InterPARES 2 opredeljuje format kot »organizacijo podatkov znotraj datoteke, ki je običajno zasnovana za lažje shranjevanje, iskanje, obdelavo, predstavitev oz. prenos po-

datkov s programsko opremo».⁵⁰ PREMIS Data Dictionary for Preservation Metadata pa definira format kot »posebne, vnaprej določene strukture za organizacijo digitalnih dokumentov ali bitnih nizov».⁵¹ Te vnaprej določene strukture vključujejo podatek o tem, kako so podatki kodirani, to je na kakšen način so biti in-

⁵⁰ Prim. Peters McLellan: *InterPARES 2 Project*, str. 2.

⁵¹ Prim. PREMIS, *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group*.

terpretirani v določenem besedilu, sliki ali zvoku. Tako so nekateri načini kodiranja sinonim za posebne oblike dokumentov, na primer kodiranje MP3se uporablja za kodiranje zvočnih datotek MP3. Lahko pa ima več formatov različne načine kodiranja, npr. golo besedilo (angl. plain text) je lahko kodirano v obliki zapisa ASCII, EBCDIC ali Unicode, vse pa imajo lahko še številne različice.

Uporabnost in razumljivost vsebine nekega dokumenta nista le zahtevi za kratkoročno časovno obdobje, temveč zahteva vseh, ki želijo do gradiva dostopati in ga uporabljati po daljšem časovnem obdobju po nastanku dokumenta. Format je odvisen od vrste dokumenta, njegove vsebine in načina uporabe ter omogoča določeni programski opremi interpretirati podatke v dokumentu oz. z drugimi podatki; je način predstavitve teh podatkov. Ustrezna strojna in programska oprema preneseta fizično raven podatkov prek predstavitvene ravni do logične, ki obvladuje različne elemente, kot so vsebina, struktura, metapodatki in kontekst. Formate zapisa moramo torej obvladovati na predstavitveni ravni. Le tako lahko pridemo do bistvenih elementov dokumenta na logični ravni.

Struktura je vezana na način, kako je dokument zapisan, to pa vključuje uporabo znakov, videza, oblike ipd. *Kontekst* je predstavljen kot dodatna informacija, ki ponazarja odnos med dokumentom in drugimi dokumenti ter aktivnost, v okviru katere je bil dokument oblikovan.⁵² Kot opozarja Bearman,⁵³ sta zlasti vsebina in kontekst tisto, s čimer je dokument opredeljen kot dokaz, t.j. umeščenost v poslovno aktivnost in upravljanje. Kontekst je lahko predstavljen z notranjimi (povezave, priloge, reference ipd.) ali zunanjimi elementi dokumenta (metapodatki), ki omogočajo razumevanje dokumenta in ga povezujejo z administrativnim in funkcionalnim okoljem ter drugimi dokumenti. Struktura in kontekst povečujeta tudi zanesljivost in avtentičnost dokumenta.

V slovenskih predpisih je »oblika zapisa« opredeljena kot tista vsebinska in tehnološka značilnost zapisa, ki določa, kako je vsebina zapisana, hranjena in prikazana v procesu hrambe.⁵⁴ ZVDAGA za obliko zapisa ob nastanku oz. v času ustvarjanja gradiva uporablja izraz »običajna digitalna oblika«,⁵⁵ z ETZ pa se ta oblika zapisa imenuje *produkcijska ob-*

lika zapisa. Običajna ali produkcijska oblika zapisa je odvisna od izvirne programske opreme, ki je omogočila nastanek oz. uporabo dokumenta.⁵⁶

Navadno oz. zelo pogosto so oblike zapisa dokumenta označene s podaljškom po zadnji piki v imenu dokumenta, npr.:

- »golo« besedilo« je najpogostejša oblika zapisa in temelji predvsem na različnih kodnih tabelah (npr. ASCII, UTF); datoteke so pogosto označene s podaljškom .txt;
- besedilni in mešani dokumenti npr. .XLS, .DOC, .WPD;
- slikovno gradivo npr. .JPEG, .PNG, .GIF, .TIF, .BMP;
- avdiovizualno gradivo npr. MPEG-2, MPEG-4;
- podatkovne zbirke, ki vključujejo našteje oblike zapisa, hkrati pa sestava zbirke zahteva dodatne posebnosti;
- spletno gradivo npr. .htm, .html.

Posebne oblike zapisa so namenjene specifičnim vrstam dokumentov ali pa so primerne za kombinacijo omenjenih vrst.

Če želimo zanesljivo ugotoviti, v kakšni obliki zapisa je določen dokument zajet, obstaja več možnosti. Prva možnost je seveda uporaba končnice imena datoteke (npr. končnica .doc pomeni obliko Microsoft Word), vendar ta način še zdaleč ni sto odstotno zanesljiv. Končnico imena datoteke lahko namreč kadarkoli zelo preprosto spremenimo. Druga možnost je, da s pomočjo ustrezne programske opreme preverimo obliko zapisa. Da dokumenta ne bi odpirali in prikazovali, to pa pomeni, da bi ob zajetu morali imeti nameščenih veliko število aplikacij za odpiranje in bi bil proces hkrati izredno časovno potraten, v dokumentu poiščemo notranje karakteristike. To je npr. zaporedje bitov, ki se enolično pojavi le v določenih oblikah zapisa, ali podpis. Te podatke lahko s posebno programsko opremo preverimo, pomagamo pa si lahko z metapodatki in registri oblik zapisa. Ti predstavljajo zbirko podatkov o oblikah zapisa, ki se centralno uporablja in posodablja, uporablja pa jo lahko različna programska oprema. Taki registri omogočajo identifikacijo oblik zapisa, validacijo, informacijo o stopnji tveganja, možnostih za migracije, karakterističnih lastnostih in načinu prikazovanja. Med najbolj znanimi omenimo:

- PRONOM,⁵⁷

⁵² Prim. ICA. *Guide for managing electronic records from an Archival Perspective*.

⁵³ Prim. Bearman: Reality and Chimeras in the Preservation of Electronic Records.

⁵⁴ ZVDAGA, 2. člen.

⁵⁵ Prav tam, 11. in 12. člen.

⁵⁶ ETZ 2.0, II. del: *Uvodne tehnološke zahteve za zajem in hrambo gradiva v digitalni obliki*, str. 41.

⁵⁷ PRONOM je spletno dostopen sistem za pridobivanje informacij o formatih datotek in programske opreme za podporo. Prvotno je bil razvit za podporo dolgoročnega hranjenja digitalnih zapisov, ki jih hrani britanski državni

- FileExr (Filerext, 2011),
- Global Digital Format Registry (GDFR, 2011) in
- TRiD.

Ta registre pogosto dopolnjujejo programska orodja in aplikativni vmesniki, s katerimi lahko registre uporabimo ali pa jih vključimo v lastne aplikacije. Omenimo lahko DROID,⁵⁸ NLNZ Metadata Extraction Tool, JHOVE in Metadata Miner Pro.

Nacionalni arhiv Velike Britanije je razvil register oblik zapisa v okviru projekta PRONOM, ki je javno dostopen spletni register in vsebuje podatke o tako rekoč vseh znanih oblikah zapisa na sveru in pripadajoče opise. Prvotno je bil razvit za podporo pri dostopu in dolgoročni hrambi digitalnih dokumentov nacionalnega arhiva. Namenjen je vsem, ki potrebujejo zanesljiv in verodostojen informacijski vir o oblikah zapisa ter programski opremi, njenih življenjskih cikli in tehničnih specifikacijah, ki jih podpira. Pozneje so v nacionalnem arhivu izdelali še aplikacijo DROID, ki avtomatsko identificira obliko zapisa dokumenta še pred dostopom (in odpiranjem) in se poveže s centralnim registrom tehničnih podatkov vseh oblik zapisa. PRONOM za natančen opis ali pa ob neznanu obliki zapisa ustvari datoteko XML s tehničnim opisom in jo pošlje v centralni register. Komunikacija z registrom PRONOM je standardizirana, tako da jo lahko vgradimo v lastno aplikacijsko rešitev za dolgoročno hrambo.

Register PRONOM vsebuje tudi seznam aplikacij, s katerimi je mogoče dokument določene oblike zapisa prikazati. Podobno odprtokodno aplikacijo JHOVE so razvili tudi na univerzi Harvard. Aplikacija dokument identificira, potrdi avtentičnost in karakterizira. Tudi JHOVE je modularne oblike, to pa pomeni, da ga lahko kdorkoli vgradi v svojo aplikacijo.

4.3 Pretvorba izvirne digitalne oblike zapisa v obliko primerno za dolgoročno hrambo

Ko dokument ustvarimo, ga zaradi različnih potreb po uporabi (npr. branja, tiskanja, urejanja, spreminjanja, pošiljanja) ali zaradi določil različnih predpisov tudi shranimo. So pa dokumenti v digitalni obliki na eni strani že sami po sebi odvisni od programske opreme, po drugi strani pa tehnološki razvoj (npr. izboljšave računalnikov, razvoj programske opreme, izvajanje več različic iste pro-

gramske opreme, obstoj različnih platform in operacijskih sistemov) povzroča oblikovanje vedno novih formatov. Ti so pogosto nezdružljivi s prejšnjimi različicami in obstoječimi platformami. Tako veliko digitalnega gradiva, ustvarjenega pred nekaj leti, ni več mogoče brati s programsko opremo, ki je na voljo danes.

Na hranjenje od programske opreme odvisnih dokumentov lahko gledamo kot na dvojni problem.⁵⁹ Prvi je ohranitev prvotne oblike in uporabnost dokumenta kadarkoli kasneje. Drugi problem pa predstavljata ustvarjanje digitalne predstavitve vsebine digitalnega dokumenta na način, ki ni odvisen od programske opreme, uporaba tega dokumenta s o programsko opremo za predstavitev prvotne oblike in vsebine dokumenta, ki od infrastrukture ni odvisna. Rothenberg⁶⁰ v poročilu za Svet za knjižnične in informacijske vire (Council on Library and Information Resources) iz začetka leta 1998 navaja, da je težave pri hranjenju digitalnih dokumentov možno reševati z različnimi metodami, kot so tiskanje digitalnih dokumentov na papir, uporaba standardov, ki ohranjajo dokumente berljive, hranjenje starejših različic programske in strojne opreme v posebnih muzejih ali pretvorba (migracija) v oblike zapisa, ki bodo dostopne s programsko opremo novjših generacij. Hkrati pa ugotavlja, da nobena navsezadnje ne ohranja digitalnih dokumentov v njihovi originalni obliki. Vsaka od omenjenih metodologij hranjenja digitalnih vsebin ima svoje prednosti in slabosti, te pa so bile in so še predmet številnih raziskav.

Nacionalna knjižnica Nizozemske (National Library of the Netherlands, Koninklijke Bibliotheek) je med najvidnejšimi zagovorniki emulacije kot najbolj učinkovite metode dolgoročnega hranjenja digitalnih dokumentov. V okviru evropskega projekta KEEP (Keeping Emulation Environments Portable), začete leta 2009, je svoj čas in denar namenila za razvoj nove programske opreme »Emulation Framework«⁶¹ s katero poskuša odgovoriti na tehnične in pravne izzive dolgoročne hrambe digitalnih dokumentov z uporabo emulacije. Programska oprema ponuja možnost za odpiranje digitalnih dokumentov in poganjanje programov v izvirnem računalniškem okolju. To uporabnikom omogoča ogled datotek na način »videti in občutiti« neodvisno od trenutne računalniške tehnologije, ki jo imajo na voljo.

arhiv, zdaj pa je PRONOM na voljo kot vir za vsakogar, ki potrebuje dostop do tovrstnih podatkov. Prim. *The technical registry PRONOM*.

⁵⁸ DROID, Digital Record Object Identification.

⁵⁹ Prim. *Methodologies for the Long-Term Preservation*.

⁶⁰ Prim. Rothenberg: *Building technological quicksand*.

⁶¹ Program je brezplačno dostopen na spletni strani <http://emuframework.sourceforge.net/>.

Bikson in Frinking⁶² pravita, da pretvorba digitalnih dokumentov lahko povzroči subtilne (občutljive) izgube konteksta in vsebine. Ugotavljala sta, da do takrat ni bilo še nobene preizkušene strategije, ki bi zagotavljala ohranitev digitalnih dokumentov v daljšem časovnem obdobju. Bila pa sta prepričana, da je za večino organizacij, ki so pričakovale, da bodo digitalni dokumenti postali neberljivi ali nedostopni zaradi zastarelosti ali nedostopnosti njihove prvotne programske opreme, pretvorba dokumentov iz izvorne v novo digitalno obliko zapisa kljub vsemu najbolj sprejemljiv pristop.

Bikson in Frinking sta že leta 1997 opozarjala, da bo treba ob morebitni odločitvi za pretvorbo iz ene digitalne oblike zapisa v drugo to početi ves čas tako, da bodo dokumenti vedno zapisani v formatu, ki ga bo mogoče v danem trenutku brati s tehnologijo, ki bo na voljo. Dodatne težave pri tem nastanejo pri uporabi vzaprtih (lasniških) oblik zapisa, saj uporaba teh ni prosta in je pogosto odvisna od licenc, posebne opreme, razširjenosti ipd.

Skladno s slovenskimi predpisi je treba, če je predpisano, da je treba gradivo, ki mora biti hranjeno več kot pet let, pretvoriti iz običajne digitalne oblike v digitalno obliko za dolgoročno hrambo.⁶³ Ko so dokumenti še v aktivni fazi življenjskega cikla in je programska oprema, s katero so bili ustvarjeni, še na voljo, lahko to storimo na dva načina:

- v trenutku nastanka ali čim prej po njem dokument že zapišemo v vnaprej določenem formatu, ki je primeren za dolgoročno hrambo;
- v kratkoročno hrambo se sprejmejo vse oblike zapisa, ki nastanejo pri poslovanju ter so navadno oblikovane za kar najučinkovitejše upravljanje in poslovanje. Šele pri zajemu v dolgoročno hrambo dobijo ustrezne oblike za tako hrambo.

V obeh primerih pa je treba v določenem trenutku izvirno digitalno obliko zapisa pretvoriti v obliko, primerno za dolgoročno hrambo. Za zanesljivo pretvorbo v digitalno obliko za dolgoročno hrambo štejemo pretvorbo, ki ohranja celovitost (iz zajetega gradiva zagotavlja reprodukcijo vseh bistvenih sestavin vsebine izvirnega gradiva glede na naravo in namen posameznih enot izvirnega gradiva) in uporabnost vsebine zajetega gradiva v običajni digitalni obliki.⁶⁴ Za hranjenje verodostojnosti gra-

diva pa mora pretvorba zagotavljati tudi avtentičnost gradiva, in sicer z zajemom ključnih vsebin izvirnega gradiva ali s strogo kontroliranim in dokumentiranim dodajanjem vsebin, ki potrjujejo enako avtentičnost zajetega gradiva, kot jo je imelo izvirno.

4.4 Lastnosti formatov, ki so primerni za dolgoročno hrambo

Formati, ki jih je stroka opredelila kot primerne za dolgoročno hrambo, morajo izpolnjevati določene kriterije. Folk in Barkstrom⁶⁵ sta identificirala šest glavnih kategorij teh kriterijev, v katere so razvrščene lastnosti teh formatov:

- preprosta dolgoročna hramba (kompaktnost, velikost, sposobnost za združitev več objektov v eno datoreko),
- preprost dostop do shranjenih dokumentov (dejanska učinkovitost, preprosto razvrščanje),
- uporabnost (razširjenost, razpoložljivost, sposobnost za integracijo programske opreme za prikazovanje podatkov s samimi podatki, preprosta uporaba, sposobnost za poimenovanje elementov datoteke),
- omogočanje formalizacije podatkov (sledljivost izvora, natančne definicije, samoopisovanje, uporaba citiranja ...),
- zagotavljanje celovitosti podatkov (verifikacija izvora, odkrivanje poškodb datoteke, popravljanje poškodovane datoteke),
- vzdrževanje in trajnost (dolgoročno vzdrževanje, primeren za različne tehnologije hrambe, stabilnost, odprtost, odprtost ...).

J. Rog⁶⁶ iz Narodne knjižnice Nizozemske je za kriterije formatov, ki so primerni za dolgoročno hrambo, definirala: odprtost, možnost razširitve in prevzeto v arhivski stroki, celovitost oz. vsestranskost, vključevanje tehničnih zaščitnih mehanizmov (npr. zaščita z geslom, zaščita pred kopiranjem, elektronski podpis), samodokumentiranost (metapodatki, vgrajen tehnični opis formata), robustnost (podpora za odkrivanje napak, stabilnost, združljivost s prejšnjimi in prihodnjimi različicami) in neodvisnost npr. od strojne in programske opreme.

Slovenski predpisi pa za splošne značilnosti oblike zapisa, ki je primerna za dolgoročno e-hrambo, določajo:⁶⁷

⁶² Prim. Bikson & Frinking: *Preserving the Present: Toward Usable Electronic Records*.

⁶³ ZVDAGA, 11. člen.

⁶⁴ IČZ 2.0-II, del: *Enotne tehnološke zahteve za zajem in hrambo gradiva v digitalni obliki*, str. 39.

⁶⁵ Folk & Barkstrom: *Attributes of File Formats for Long-Term Preservation*, str. 3.

⁶⁶ Prim. Rog & Wijk: *Evaluating File Formats for Long-term Preservation*.

⁶⁷ ZVDAGA, 14. člen.

Št.	Lastnost
1	Zagotavlja obratitev vsebine gradiva, tako da pomeni urejeno celoto vseh potrebnih podatkov in povezav med njimi.
2	Je široko priznana in uveljavljena oziroma v uporabi, uporaba podprta s strojno in programsko opremo, uveljavljeno na trgu.
3	Neposredno uporabna za reprodukcijo vsebine ali enostavno pretvorjena v obliko, ki je neposredno uporabna.
4	Omogoča samodejno pretvorbo iz najpogostejše uporabljanih izvornih oblik zapisa s samodejno zažupavo in poročanjem o nepredvidenih dogodkih oziroma napakah pri pretvorbi.
5	Je neodvisna od posamezne programske ali strojne opreme oziroma okolja
6	Glede na stanje stroke obstaja velika stopnja verjetnosti, da zagotavlja varno hrambo več kot pet let.
7	Omogoča po današnjih strokovnih predvidevanjih po tem obdobju pretvorbo v novo, takrat določeno obliko zapisa za dolgoročno hrambo.
8	Temelji na mednarodnem, državnem ali splošno priznanem in praviloma odprtem standardu, če obstaja.

Tabela 1: Splošne lastnosti formata za dolgoročno hrambo, kot jih določajo slovenski predpisi (U.V.D. 1G, 14. člen).

Ker UVDAGA določa zgolj splošne lastnosti oblike zapisa, ki je primerna za dolgoročno digitalno hrambo (glej Tabelo 1), je Arhiv RS z ETZ za posamezne tipe gradiva konkretnije navedel oz. priporočil tiste oblike zapisa, ki te lastnosti imajo in so bile, ko so bile sprejete, najpogostejše v uporabi.⁶⁸

ETZ kot enega izmed priporočenih formatov za dolgoročno hrambo besedilnih in mešanih dokumentov priporoča PDF/A (ISO 19005).

V skupino »besedilni in mešani dokumenti« so dijo dokumenti, ki jih ustvarjamo s sodobnimi t. i. zbirkami pisarniških programov, s katerimi urejamo npr. besedila, preglednice in predstavitev. Značilen primer zbirke pisarniških programov je Microsoftov Office, ki je zdaj najbolj razširjena zbirka pisarniških programov na svetu in predstavlja »de facto« standard,⁶⁹ čeprav se ji manjša tržni delež zaradi odprtokodnih alternativ, kot sta OpenOffice.org in StarOffice. Pisarniška zbirka programov Microsoft Office vključuje:

- Word – urejevalnik besedil,
- Excel – program za delo s preglednicami,
- Outlook – urejevalnik opravil, integriran s koledarjem in odjemalcem elektronske pošte ter knjigo naslovov in
- Power Point – program za izdelavo predstavitev, sestavljenih iz besedil, slik in animacij.

Če se namesto na tip osredotočimo na vsebino gradiva, gre najpogostejše za upravno-poslovno dokumentacijo v poslovnem okolju, spisovno gradivo, ki ga vodimo v evidenci dokumentarnega gradiva, in specifične vrste gradiva, kakršne so finančna, računovodska in knjigovodska dokumentacija, projektna in tehnična dokumentacija, šolska in medicinska dokumentacija itd.

Gre torej za dokumente z oblikovanimi besedili (oblikovni učinki, številčenje, naštevanje, tabele) in so bile za njih izdelane posebne oblike zapisa. Te poleg osnovnega oblikovanja (različne pisave, velikosti, učinki ipd.) omogočajo vključevanje matematičnih formul, tabel, slik in drugih elementov.

4.5 Raziskovalna vprašanja

S slovenskimi predpisi (ETZ) je torej priporočen format za dolgoročno hrambo besedilnih in mešanih dokumentov PDF/A.⁷⁰ Z metodo primerjanja⁷¹ primarnih in sekundarnih virov smo preučili, ali format PDF/A izpolnjuje vse lastnosti, ki jih določa ZVDAGA za formate, ki so primerni za dolgoročno hrambo (glej Tabelo 1).

Zastavili smo si ta vprašanja:

1. ali pretvorba besedilnega dokumenta iz produkcijske oblike v format PDF/A vpliva na vsebino in obliko dokumenta?

⁶⁸ ETZ 2.0-1, del: Uvodna poglavja, str. 12.

⁶⁹ »De facto« standardi (imenujemo jih tudi industrijski) nastajajo s prakso vodilnih na trgu na določenem področju. De facto standardi lahko postanejo uradni standarde, če se vodilni na trgu odločijo, da je vlaganje v trg bolj koristno kot povečevanje njihovega tržnega deleža.

⁷⁰ PDF/A je tudi ISO 19005 standard.

⁷¹ Metoda primerjanja (komparativna metoda): preučevanje na ravni primerjanja dejstev, odnosov, procesov za odkrivanje podobnosti in razlik. Primarni viri vsebujejo nova, izvirna spoznanja ali novo izvirno razlago sicer že znanih spoznanj – v to skupino sodijo knjige, članki v strokovnih in znanstvenih revijah, zborniki, magistrska dela, doktorske disertacije. Sekundarni viri pa vsebujejo informacije iz primarnih virov – v to skupino sodijo učbeniki, priročniki, enciklopedije, slovarji.

2. ali ima standard PDF/A vse lastnosti, ki jih slovenski predpisi določajo za oblike zapisa, ki so primerne za dolgoročno hrambo besedilnih in mešanih dokumentov?
3. ali je pretvorba dokumenta iz ene digitalne v drugo digitalno obliko zadostna strategija za njegovo dolgoročno hranjenje?

5. PDF/A – format za dolgoročno hrambo besedilnih in mešanih dokumentov

5.1 Zgodovinski razvoj formata PDF/A⁷²

Leta 1990 se je pojavila oblika zapisa PDF (Portable Document Format), ki je široko uporabljena v vsem svetu zlasti zaradi vpetosti v področje založništva, brezplačnega programa za branje dokumentov (Adobe Reader) in možnosti za uporabo programa na različnih platformah (Windows, Linux ipd.). Je vsesplošno integrirana v poslovna okolja, navzoča na svetovnem spletu in z njo so dobro seznanjeni številni uporabniki. Oblika zapisa PDF je, kot kaže, za številne edina izbira in večinoma ustreza zahtevam ustvarjalcev dokumentov, uporabnikom in arhivistom. Specifikacija PDF je javno objavljena in trenutna različica je, tako kot tudi vse predhodne različice na voljo na svetovnem spletu. PDF je oblika zapisa, ki ni namenjen urejanju vsebine, temveč prikazovanju in hranjenju »videza in občutka« ter je neodvisna od platforme. Najpogostejše ga ustvarimo, ko končno oblikovan dokument, na primer besedilo, napisano z urejevalnikom besedil Microsoft Word, pretvorimo v obliko zapisa PDF.

V osnovi PDF ni primerna oblika zapisa za arhiviranje. Lahko vsebuje kodirano vsebino, lahko vsebuje vstavljene druge zapise ali datoteke in ni nujno samozadostna. Dokumenti PDF namreč lahko vključujejo na primer pisavo (font) ali sliko, ki ni vključena v dokument. Ameriško računalniško podjetje Adobe Systems Incorporated nadzira razvoj oblike zapisa in ni obvezano, da dolgoročno obliko zapisa ohranja in razvija ali celo objavlja specifikacijo prihodnjih različic. Poleg tega lahko oblika zapisa vsebuje elemente, ki so nekompatibilni s predhodnimi različicami oblike zapisa.⁷³ Težava je v tem, da je specifikacija oblike zapisa PDF

javna, to pa omogoča različnim razvijalcem programske opreme izdelavo programov, ki izdelajo dokument v obliki zapisa PDF (med drugim tudi najnovejši paket Microsoft Office 2007/2010 vključuje možnost pretvorbe v obliko zapisa PDF). Ti razvijalci lahko kljub specifikaciji oblikujejo zapis PDF na različne načine, to pa včasih povzroča manjše nekompatibilnosti določenih dokumentov, zapisanih v obliki zapisa PDF. Razvijalci lahko vključijo elemente, ki jih Adobe ni predvidel. Tako mora Adobe, ki se je zavezal, da bo omogočal branje vseh dokumentov PDF, skrbeti tudi za vse dokumente PDF, ki niso bili izdelani z njihovimi orodji. Če bi bila specifikacija PDF zaprta, bi se podjetje Adobe takim težavam verjetno izognilo.

V portfelj standardov PDF zdaj sodi več tipov: PDF, PDF/A, PDF/E, PDF/UA, PDF Healthcare in PDF/X. V nadaljevanju bomo obravnavali le format PDF/A, ki je skladno s slovenskimi predpisi format, primeren za dolgoročno hrambo besedilnih in mešanih dokumentov.

PDF/A

PDF/A (družina standardov ISO 19005) je standardizirana verzija oblike zapisa PDF. Od PDF se razlikuje v tem, da so odstranjene značilnosti, ki so neprimerne za dolgoročno hrambo kot npr. povezovanje (sklicevanje) na pisave (fonte) je nadomeščeno z zahtevo po vključenosti uporabljenih fontov v sam dokument.

Standard PDF/A ne določa strategije hrambe ali ciljev sistema digitalne hrambe. Identificira pa »profil« za digitalne dokumente, ki bo zagotavljal, da bodo lahko dokumenti za ves čas njihove hrambe reproducirani vedno na natančno enak način. Ključni element reprodukcije je zahteva, da so PDF/A dokumenti 100-odstotno neodvisni. Vse informacije, ki so potrebne za prikaz dokumentov vedno na enak način, morajo biti vgrajene v datoteko. To vključuje (ni pa omejeno na) informacijo o vsebini (besedilo, rastrske slike in vektorske grafike), pisavah in barvah. Dokument PDF/A ne sme biti odvisen od podarkov iz zunanjih virov (npr. fontov programa in hiperpovezav). Zahteve ISO standarda za programsko opremo, ki prikazuje format PDF/A so, da vključuje smernice za upravljanje z barvami, podporo za vgrajene pisave (fonte) in uporabniške vmesnike za branje vgrajenih opomb (npr. podčrtavanje ali označevanje besedila).

⁷² Glej npr. <http://www.pdfa.org/2011/08/pdfa-a-new-standard-for-long-term-archiving/>.

⁷³ Primer iz prakse kaže, da se napake lahko pojavijo tudi pri mogočih razvijalcih programske opreme, kot je Adobe. Npr. z različico 7 se je pojavila napaka pri odpiranju določenih dokumentov, ki so bili izdelani s programsko opremo drugih proizvajalcev ali produktom Adobe Distiller 3.0.2. Napaka je bila v napačni interpretaciji oziroma prikazu pisave oziroma sesutju programa za branje dokumentov PDF Adobe Reader.

<i>Leto izdaje</i>	<i>PDF</i>	<i>Adobe Acrobat</i>	<i>Format</i>	<i>ISO</i>
1993	PDF 1.0	Acrobat 1.0	PDF	
1994	PDF 1.1	Acrobat 2.0	PDF	
1996	PDF 1.2	Acrobat 3.0	PDF	
1999	PDF 1.3	Acrobat 4.0	PDF	
2001	PDF 1.4	Acrobat 5.0	PDF/A-1a PDF/A-1b	ISO 19005-1:2005
2003	PDF 1.5	Acrobat 6.0	PDF/A	
2005	PDF 1.6	Acrobat 7.0	PDF/A	
2006	PDF 1.7	Acrobat 8.0	PDF/A	
2006	PDF 1.7	Acrobat 82	PDF/A	
2008	PDF 1.7	Adobe Extension Level 3 / Acrobat 9.0	PDF/A	ISO 32000-1:2008
2009	PDF 1.7	Adobe Extension Level 5 / Acrobat 9.0	PDF/A	
2010	PDF 1.7	Adobe Acrobat X (version 10.0)	PDF/A-2a PDF/A-2b	ISO 19005-2010

Tabela 2: Zgodovinski pregled verzij PDF in PDF/A

(Prirejeno iz vira http://en.wikipedia.org/wiki/Portable_Document_Format#cite_note-pdf-ref-9#cite_note-pdf-ref-9, najdeno dne 10.11.2011)

Standard ISO za PDF/A prepoveduje avdio in video vsebine, zunanje povezave, ukaze JavaScript in ukaze za sprožanje ukazov in odpiranje drugih dokumentov, vključevanje drugih zapisov in enkripcijo.

Nadalje zahteva, da so vse uporabljene pisave (fonti) vključene v sam dokument in da so hkrati univerzalne ter brez pravnih omejitev za njihovo uporabo. Zahteva tudi, da je barvna shema določena v obliki, neodvisni od strojne in programske opreme, metapodatki pa morajo biti zapisani v obliki zapisa XMP.⁷⁴

Oblika zapisa PDF/A ima dva prilagojena nivoja in verziji:

- PDF/A-1 (ISO 19005-1):

Temelji na verziji PDF 1.4 od Adobe Systems Inc. (vključena v Adobe Acrobat 5 in poznejše različice) in je definirana kot ISO 19005-1:2005. Objavljena je bila 1. oktobra 2005 z imenom »Document Management - Electronic document file format for long term preservation - Part 1: Use of PDF 1.4 (PDF/A-1)«.

V to družino standarda sodita dve podrazličici:

- PDF/A-1a (ISO 19005-1 nivo A skladnosti): zagotavlja hranjenje enake logične strukture, konteksta in vsebine dokumenta kot izvorna oblika.

- PDF/A-1b (ISO 19005-1 nivo B skladnosti): zagotavlja hranjenje enake vizualnosti dokumenta kot izvorna oblika.

- PDF/A-2 (ISO 19005-2:2011):

Temelji na verziji PDF 1.7 od Adobe Systems Inc., ki je standard ISO 32000-1. Objavljena je bila 20. junija 2011 kot ISO 19005-2:2011 z imenom »Document management - Electronic document file format for long-term preservation - Part 2: Use of ISO 32000-1 (PDF/A-2)«. PDF/A-2 je torej prav zdaj nastala različica in zato še ni zelo splošno uporabljena.

Adobe Systems Inc. je specifikacijo PDF večkrat spremenil in še vedno nadaljuje razvoj v obliki novih različic Adobe Acrobat. Obstaja deset različic PDF s pripadajočimi različicami Adobe Acrobat (glej Tabelo 2).

5.2. Značilnosti PDF/A

Pomembna informacija je, da so vse različice združljive s prejšnjimi (npr. različica 1.7 vključuje vse v Adobe PDF specifikacijah dokumentirane funkcionalnosti za različice od 1.0 do 1.6). Ker je specifikacija javna (odprta oblika zapisa), podjetje Adobe kot avtor in lastnik oblike zapisa (lastniška oblika zapisa) nima monopola nad obliko zapisa, vendar pa je zaradi zahtevnosti standarda ter občasnih sprememb in dopolnitev podpora obliki zapisa odvisna od podjetja Adobe. Leta 2008 je PDF 1.7 postal ISO standard 32000-1:2008.

⁷⁴ XMP (Extensible Metadata Platform) specifikacija, temelji na standardu XML in omogoča uporabo shem za opis metapodatkov.

Osnovne značilnosti PDF/A-1 so:⁷⁵

- enkripcija je prepovedana;
- JavaScript in ukazi za proženje ukazov so prepovedani;
- avdio in video gradivo sta prepovedani;
- obvezna je uporaba metapodatkov, ki slonijo na standardih;
- barvne sheme morajo biti določene v obliki, neodvisni od strojne in programske opreme;
- vse pisave (fonti), uporabljene v izvornem digitalnem dokumentu, morajo biti vključene v dokument. To velja tudi za t. i. standardne pisave PostScript, kot sta Times ali Helvetica;
- opombe, predstavljene kot govor ali video, niso dovoljene;
- sklicevanje na zunanje vsebine ni dovoljeno;
- akcije, kot npr. zagon filmskih ali zvočnih datotek, pošiljanje ali brisanje formularjev niso dovoljene;
- stiskanje LZW in JPEG2000 (kompresija) ni dovoljeno;
- transparentnost in uporaba slojev nista dovoljeni;
- vgrajeni dokumenti (datoteke) so prepovedani v PDF/A-1.

Kaj prinaša novega PDF/A-2:

PDF/A-2 ima dodane nekatere nove funkcije glede na specifikacije različic PDF 1.5, 1.6 in 1.7 in je združljiv z vsemi prejšnjimi različicami. Torej so dokumenti, ki so skladni s PDF/A-1, skladni tudi s PDF/A-2. Vendar pa dokumenti, ki so skladni s PDF/A-2, niso nujno skladni s PDF/A-1 (Wikipedia, 2011b). Novosti glede na PDF/A-1 so:

- JPEG 2000 stiskanje je dovoljeno;
- podprt je digitalni podpis, skladen s PAdES (PDF Advanced Electronic Signatures);
- transparentnost in uporaba slojev sta dovoljeni;
- vgraditev datotek PDF/A je dovoljena;
- nivo podpore kodni tabeli Unicode;
- odprte oblike zapisa (angl. *OpenType fonts*);
- novi tipi zaznamkov (angl. *New annotation types*), npr. vodni žig (angl. *Watermark*), zaznamki urejanja (angl. *Redact annotations*).

Kaj ni dovoljeno v PDF/A-2:⁷⁶

Prepovedane so vgrajene dinamične vsebine v datoteke PDF/A. Obstaja nekaj področij, na katerih dinamične vsebine igrajo pomembno vlogo:

- multimedijske lastnosti: zaznamki, ki omogo-

čajo vključevanje filmov, avdio zapisov ali modelov 3D v PDF datoreke;

- obrazci, zasnovani na formatu XFA – nova arhitektura, ki jo je razvilo podjetje Adobe za podporo zahtevnim interaktivnim obrazcem.

Odnos do akcijskih skript (ActionScripts) in Java skript je še vedno enak kot v PDF/A-1. Java skripte so popolnoma prepovedane v PDF/A-2, dovoljeno je samo zelo majhno število ukazov akcijskih skript, in sicer le tisti, za katere ne obstaja tveganje, da lahko spremenijo vsebino arhiviranega (shranjenega) dokumenta PDF.

5.3 Pregled razširjenosti PDF/A v mednarodnih inštitucijah

V nadaljevanju nas je zanimalo, v kakšni meri je format PDF/A (vključno z različnimi verzijami) priznan oz. uporabljen v različnih mednarodnih inštitucijah. Organizacija AIIM⁷⁷ je v tržni raziskavi⁷⁸ ugotovila, da 90% organizacij, ki so bile zajete v študiji (198 članic AIIM), uporablja PDF za dolgoročno hrambo skeniranih dokumentov, 89% organizacij pa za potrebe distribucije in hrambe pretvarja dokumente pisarniških zbirk v format PDF. Vse organizacije pa še vedno uporabljajo papir.

Organizacije, zajete v študiji AIIM, so napovedale, da bodo v naslednjih petih letih zmanjšale porabo papirja na 77% in povečale uporabo PDF na 93%.

V drugi raziskavi,⁷⁹ ki se je posvetila uporabi PDF/A v Evropi, je kompetentni sektor PDF/A (PDF/A Competence Center) ugotovil, da jih od 400 vprašanih organizacij 16% že uporablja PDF/A, v naslednjih 12 mesecih pa ga namerava začeti uporabljati že 50% vprašanih.

Po podatkih, ki jih je objavila družba Adobe Systems Incorporated (Adobe Acrobat, 2010), so vladne organizacije med prvimi uradno definirale uporabo standarda PDF/A. Tako na primer:

- NARA (National Archives and Records Administration) *dovoljuje uporabo* PDF/A pri sprejemanju dokumentov (submission). PDF/A *dovoljujejo* ali *priporočajo* nacionalni arhivi ali knjižnice v Nemčiji, Franciji, Italiji, Švedskem, Nizozemskem, Avstriji, Avstraliji in Norveškem;

⁷⁵ Najdeno na naslovu: <http://en.wikipedia.org/wiki/PDF/A> (15.11.2011).

⁷⁶ Prim. Drümmer: *What is PDF/A-2?*

⁷⁷ AIIM (Association for Information and Image Management) – Mednarodno združenje za upravljanje informacij.

⁷⁸ Prim. Frappaolo: *AIIM Market IQ study, Content Creation and Delivery*.

⁷⁹ Prim. *Nearly all archiving projects use PDF/A*.

Institucija	Obvezna uporaba (najvišja stopnja zaupanja)	Priporočena uporaba (visoka stopnja zaupanja)	Dovoljena uporaba (srednja stopnja zaupanja)
Arhivi:			
Arhiv RS (ETZ 2.0)		✓	
Florida Digital Archive		✓	
Library and Archives Canada		✓	✓
National Archives of Norway		✓	
National Archives of Sweden			✓
SWISS Federal Archives (SFA)		✓	
U. S. National Archives and Records Administration (NARA)			✓
Knjižnice, muzeji:			
Austrian National Library		✓	
Bibliothèque nationale de France		✓	
German National Library		✓	
Library and Archives Canada		✓	✓
Museum of London		✓	
University of Texas Libraries		✓	
Organizacije javnega sektorja (Government agencies):			
Brazilian federal legislature	✓		
Dutch government	✓		
Franch	✓		
German government (SAGA v4)		✓	
Norway		✓	
Switzerland		✓	
Spanish (Extremadura)	✓		
Preostalo:			
European Commission (MoReq)		✓	
Organization for the Promotion of Automated Accounting		✓	
United States Patent and Trademarks Office (USPTO)	✓		
Virginija - visoke šole za podiplomski študij	✓		

Tabela 3: Seznam mednarodnih institucij, ki vključujejo PDF/A v nabor oblik zapisa za dolgoročno hrambo in stopnja uporabe (Spletne strani institucij).⁸⁰

http://en.wikipedia.org/wiki/OpenDocument_adoption,

<http://digitalpreservation.gov/formats/jdd/jdd000125.shtml>, <http://www.adobe.com/enterprise/standards/pdfa/>)

⁸⁰ Najdeno 18.11.2011 na naslovih:

<http://www.adobe.com/enterprise/pdfs/pdfaforAcrobat.pdf>

<http://www.collectionscanada.gc.ca/digital-initiatives/012018-2210-e.html>

http://www.arhiv.gov.si/fileadmin/arhiv.gov.si/pageuploads/zakonodaja/ETZ_2_0_-_1_del_razlicica_2.0.pdf

http://www.cha.ec/wiki/index.php/Norway#Country_Profile_for_Norway

<http://www.collectionscanada.gc.ca/digital-initiatives/012018-2210-e.html>

<http://www.culture.fr/culture/mrt/numerisation/fr/technique/documents/>

[cahier_charges_numerisation_presse_annexes.doc](#)

http://www.museumoflondonarchaeology.org.uk/English/ArchiveResearch/DeposResource/Standards_2.htm

<http://www.cha.ec/wiki/index.php/Norway>

http://repositories.lib.utexas.edu/recommended_file_formats

<http://www.d-nb.de/eng/netzpub/ablief/np-technik.htm>

- evropska komisija je v specifikaciji MoReq (Model Requirements for the Management of Electronic Documents and Records) format PDF/A *priporočila* najmanj za skenirane dokumente in dolgoročno hrambo;
- sodišča ZDA so kot ena prvih, ki so privzela Format PDF/A kot *obvezen* v primeru sprejemanja uradnih pritožb, stečajnih napovedi, odvetniških zapisov in vlog. Ob prejemu te dokumente pretvorijo v PDF/A in časovno žigosajo ter podpišejo pred nadaljnjo uporabo;
- države, kot so Nemčija, Francija in Nizozemska, *priporočajo* PDF/A na nacionalni ravni, kadar je treba ohraniti administrativne dokumente s fiksno, nespremenljivo vsebino.

S pomočjo dostopnih virov smo raziskali, kakšna je stopnja zahtevnosti uporabe formata PDF/A v različnih mednarodnih inštitucijah, kot so arhivi, knjižnice, muzeji, univerze in še nekatere druge (opazovane inštitucije). Zahteve po uporabljanju formata PDF/A smo razdelili tako: obvezna, priporočljiva ali dovoljena uporaba (glej Tabelo 3).

Iz Tabele 3 je vidno, da so mednarodne inštitucije previdne, saj jih PDF/A večina le priporoča (16 inštitucij od 24 opazovanih, v glavnem arhivi, knjižnice in javni sektor), kot obveznega ga določa 6 od 24 opazovanih inštitucij (javni sektor, noben arhiv ali knjižnica), štiri inštitucije pa zgolj dovoljujejo uporabo formata PDF/A za dolgoročno hrambo gradiva v digitalni obliki (arhivi in knjižnice).

5.4 Pretvorba besedilnega dokumenta iz izvorne v PDF/A obliko zapisa

Pri oblikovanju vsebine besedilnega dokumenta (npr. ustvarjenega z MS Word ali MS Excel) v njegovi izvorni obliki lahko uporabimo npr.:

- formule,
- makroje,
- tabele,
- grafe, ki so izdelani na podlagi podatkov v tabeli,
- različne tipe pisav (fontov), ki lahko vključujejo šumnike oz. druge posebne znake,
- varnostne vsebine (npr. e-podpis, časovni žig),
- slike in povezave na zunanje dokumente (vgrajeni dokumenti).

V času ustvarjanja dokumenta kor rudi kasneje je pomembna tudi oblika dokumenta, npr.:

- prelom strani,
- izvorna pisava (font),
- videz npr. logotip ali slike,

- prelomi vrstic,
- barve,
- označevanje in številčenje (slogi in oblikovanje naslovov oz. odstavkov),
- formati posameznih tipov podatkov (npr. decimalne vejice pri številkah, datum, čas, valuta),
- postavitev in videz glave/noge,
- posebni učinki, npr. poudarjeno besedilo (bold), poševna pisava (italic), podčrtano besedilo.

Če upoštevamo določila slovenskih predpisov, je treba dokumente, kolikor je hramba predpisana za več kot pet let, pretvoriti v obliko zapisa za dolgoročno hrambo. Besedilne dokumente torej v format PDF/A.

V zapisu PDF/A-1a lahko zapišemo takò besedilo kot grafične elemente in ohranimo pisavo, strukturo, obliko, barve in slike. Ohranja prikaz oziroma videz in omogoča uporabo metapodatkov. Prav tako omogoča izvedbo operacij v besedilu, kot sta iskanje po besedilu in kopiranje v druge dokumente, če seveda besedilo v samem dokumentu PDF/A ni vključeno kot slika, tj. posnetek besedila, to pa je po specifikaciji PDF tudi mogoče.

Odgovor na vprašanje lahko najdemo takò v javno dostopnih specifikacijah formata PDF/A-1 (glej poglavje 5.2. Značilnosti PDF/A) kot tudi v različnih drugih dostopnih virih, npr. rubrikah »odgovori na vprašanja« kompetenčnega centra PDF/A⁸¹ ali objavljenih člankih in brošurah⁸². Izkušnje pri pretvorbi besedilnih dokumentov, ustvarjenih s programoma Word in Excel iz pisarniške zbirke programov MS Office, je opisal tudi Bergmann iz podjetja Siemens. Bergmann navaja, da so imeli pri pretvorbi projektne dokumentacije, ki je bila v večjem delu tudi v obliki datotek v Wordu in Excelu, veliko težav. Razlogi zanje so bili pri različnih dokumentih različni, to pa se je kazalo v tem, da se je proces pretvorbe ustavil in je bila potrebna ročna intervencija sistemskega administratorja. Identificiral je nekaj skupnih težav, ki se nanašajo na makroje, skripte Visual Basic, datoteke samo za branje (read-only), preštevilčenje iz Word datotek v datoteke z več kot 500 stranmi, na datoteke z več kot 100 MB in več kot 5000 stranmi. V večini primerov so težave pri pretvarjanju izhajale iz enega od teh razlogov. Navaja, da zahteva PDF/A po neuporabi makrojev in skript ni ustrezna izbira za projektno dokumentacijo, saj je z njihovo pomočjo lažje ustvarjati in delati z dokumenti, preden so preneseni v njihov sistem PLM Cadim (sistem za

⁸¹ Prim. *How to answer questions in PDF documents?*

⁸² Prim. Drümmer & Seggern: *PDF/A in a Nutshell: Long Term Archiving with PDF*.

upravljanje specifikacij produktov, v katerega prihajajo podatki iz različnih sistemov kot CAD, CAE, SAP). Problem so sicer poskusili rešiti, vendar kot arhivski format za tehnično dokumentacijo še vedno uporabljajo format TIFF G4. Kreirajo dokumente PDF in jih pretvarjajo v TIFF G4, načrtujejo pa vpeljavo formata PDF/A, ki bo vključeval nove funkcionalnosti.

Lasten primer 1:

Ne glede na to, da v opisih formata PDF/A zveemo, da vgraditev zunanjih datotek v datoteko ni dovoljena, je dejstvo, da tak način ustvarjanja vsebin pri vsakdanjem delu s pridom uporabljamo. Takšen primer je lahko analitično poročilo, ki temelji na podatkih, pridobljenih s pomočjo dodatnih analiz ali izračunov, zapisanih v samostojnih dokumentih. Pogosto zato, da bi dosegli preglednost takšnega poročila, hkrati pa zagotovili dostop do vseh podatkov, vgradimo te samostojne datoteke v poročilo (glej Sliko 3). V konkretnem primeru smo v dokument vgradili excelovo datoteko (datoteka .xls) s funkcijo:

- * vstavi predmet
- * ustvari iz datoteke
- * prebrskaj
- * prikaži kot ikono.

Že iz Slike 3 vidimo, da lahko vgrajeno datoteko poleg tega, da jo izrežemo in kopiramo, tudi odpremo in urejamo. Pri tem ni treba, da je vgrajena datoteka v našem računalniku.

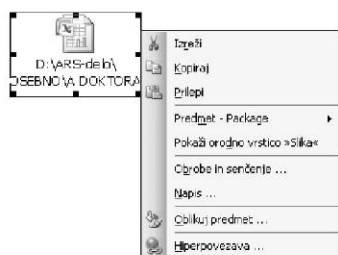
Datoteko z vgrajeno drugo datoteko smo pretvorili v format PDF/A-1. (glej Sliko 4)

Iz Slike 4 je vidno, da po pretvorbi ostane vgrajena datoteka samo še kot slika (angl. *image*), njena vsebina pa je nedostopna. Takoj ko tako datoteko pretvorimo v format PDF/A, sicer ohranimo njen videz, celotne vsebine pa ne več. Izgubimo pa tudi možnost za ponovno uporabo podatkov kasneje, saj je vsebina vgrajene datoteke izgubljena. To pa dolgoročno pomeni tudi to, da ne bo več mogoče razumeti, kako so podatki v osnovni datoteki nastali. Podobno se zgodi tudi, če imamo datoteko .xls, v katero smo zapisali niz podatkov, ki smo jih obdelali z različnimi formulami, nato pa končne rezultate prikazali v obliki grafa. Po pretvorbi datoteke .xls v format PDF/A izgubimo možnost ponovne uporabe podatkov.

6 – Stran

Test

Primer vstavljene datoteke .xls (ustvarjena z MS Excel):



Slika 3: Primer vstavljene datoteke kot slike (ikone) v drugo datoteko

6 – Stran 7

Test

Primer vstavljene datoteke .xls (ustvarjena z MS Excel):



Slika 4: Datoteka z vgrajeno drugo datoteko po pretvorbi v PDF/A-1

Pri pretvorbi dokumenta v format PDF/A-1a nas je najprej zanimalo, kakšen vpliv ima ta pretvorba na vsebino in obliko dokumenta (raziskovalno vprašanje št. 1).

Odgovor na vprašanje št. 1:

S pomočjo primerjanja primarnih in sekundarnih virov, lastnega primera in drugih dobrih praks lahko sklenemo, da pretvorba dokumenta iz njegove izvorne digitalne oblike v PDF/A lahko vpliva na obliko in predvsem vsebino dokumenta.

5.5 Ali PDF/A izpolnjuje kriterije slovenskih predpisov?

V nadaljevanju smo poskusili s pomočjo primerjanja primarnih in sekundarnih virov ter prej opisanih primerov za vsako posamezno lastnost, ki jo slovenska zakonodaja določa za formate, primerne za dolgoročno hrambo, ugotoviti, ali ga format PDF/A izpolnjuje ali ne (raziskovalno vprašanje št. 2).

Odgovor na vprašanje št. 2:

S pomočjo primerjanja primarnih in sekundarnih virov, lastnega primera in drugih dobrih praks lahko zaključimo, da za skladaj format PDF/A ne izpolnjuje vseh lastnosti formata za dolgoročno hrambo gradiva v digitalni obliki, kot jih določajo slovenski predpisi (glej Tabelo 1).

5.6 Ali je pretvorba dokumenta iz ene digitalne oblike v drugo dovolj primerna strategija za dolgoročno hranjenje?

Že pri primeru pretvorbe besedilnih dokumentov (npr. ustvarjenih z Wordom ali Excelom) v format PDF/A smo lahko videli, da niso zagotovljena vsa načela varne hrambe, kot jih opredeljuje stroka, npr. ISO 15489-1, International Council on Archives (ICA) in ZVDAGA.

Odgovor na vprašanje št. 3:

Na podlagi metode primerjanja (komparativne metode) primarnih in sekundarnih virov ter preučitve lastnega primera pretvorbe dokumenta iz ene digitalne oblike v drugo ocenjujemo, da strategija pretvorbe digitalnega dokumenta iz ene digitalne oblike v drugo ni zadostna strategija za dolgoročno hranjenje digitalnih dokumentov.

1. ISO 15489-1: dolgoročno hrambo definira kot dolgoročno shranjevanje digitalne informacije brez napak z možnostjo ponovne pridobitve (angl. *retrieval*) in interpretacije (angl. *interpretation*) za ves čas hrambe te informacije. Pri tem »ponovna pridobitev« pomeni pridobitev potrebnih, dolgoročno

hranjenih digitalnih dokumentov brez napak in morebitnih poškodb. Pri pretvorbi dokumentov z vgrajenimi dokumenti, uporabljenimi makroji in formulami v format PDF/A ta zahteva ni izpolnjena.

2. International Council on Archives (ICA): izpostavlja načelo morebitne ponovne uporabe. To pravi, da so dokumenti *potencialno ponovno uporabni*, če je iz njih mogoče izluščiti informacije ali kako drugače omogočiti dokumentu, da sodeluje s sodobnimi informacijskimi sistemi. Zahtevo po ponovni uporabi je sicer mogoče tudi zanemariti, saj je v večini primerov ne bomo potrebovali, v določenih primerih pa jo bo vendarle zaželeno doseči. Na tak način sta navadno zagotovljeni tudi dostopnost in razumljivost dokumentov. Glede na pregledane vire so stališča v stroki različna. Če pogledamo gradivo, ki je po svojem statusu arhivsko gradivo, verjetno takšna zahteva ni nujna. Če pa pogledamo gradivo, ki je po svojem statusu trajno gradivo (in ni arhivsko), bomo verjetno našli primere, za katere bi bilo izpolnjevanje tega načela smiselno.

3. ZVDAGA: s stališča izpolnjevanja načel varne hrambe, kot jih določa ZVDAGA, lahko prav tako ob obravnavanem primeru pretvorbe besedilnih dokumentov v format PDF/A ugotovimo, da načela niso v popolnosti izpolnjena:

- **uporabnost** (*Zmožnost za reprodukcijo in primernost reprodukcije za uporabo ves čas trajanja hrambe (npr. branje, tiskanje dokumenta).* V našem primeru pretvorbe ne moremo zagotoviti zmožnost za reprodukcijo vsebine dokumentov, ki so v vgrajenem dokumentu pretvorjenega dokumenta;

- **celovitost** (*Zagotovljene morajo biti pravilnost, nespremenljivost in popolnost gradiva oziroma reprodukcije njegove vsebine, urejenost gradiva oziroma njegove vsebine ter dokazljivost izvora gradiva (provenience) ves čas njegove hrambe).* Zatakne se pri zagotavljanju »popolnosti gradiva«. V slovarju slovenskega jezika najdemo obrazložitev besede popolnost kot: »popolnost -i (un) ž (o) lastnost, značilnost popolnega: popolnost seznama / nar. ekspr.: doseči popolnost; ki zajema kaj v celoti, obsega vse sestavne dele celote, enote, ki je po svojih lastnostih, značilnostih brez napak, pomanjkljivosti: popoln sistem«. V našem primeru pretvorbe torej popolnosti gradiva v tem pomenu nismo zagotovili;

- **dostopnost** (*Gradivo oziroma reprodukcija njegove vsebine mora biti ves čas trajanja hrambe zavarovano pred izgubo ali okrnitvijo celovitosti ter dostopna tistim, ki imajo za to ustrezna pooblastila).* V našem primeru pretvorbe nismo zagotovili zavarovanja gradiva pred izgubo, saj se je s pretvorbo datoteke, v katero je bila vgrajena druga datoteka (enako pa velja tudi

za uporabo makrojev in formul), vsebina vgrajenih darotek izgubila;

- **avtentičnost** (dokazljiva povezanost reproducirane vsebine z vsebino izvirnega gradiva oziroma izvorom

tega gradiva). Po prervorbi v osnovni daroteki, ki je bila predmet prervorbe, sicer vidimo vgrajeno datoteko kot sliko, njena vsebina pa je izgubljena oz. se ob prervorbi ni reproducirala.

Št.	Lastnost	Ugotovitev skladnosti ⁸³	Utemeljitev
1	Zagotavlja ohranitev vsebine gradiva tako, da pomeni urejeno celoto vseh potrebnih podatkov in povezav med njimi.	NE	Če v besedilni dokument (npr. Word, Excel) vgradimo drug dokument (s funkcijo »vstavi predmet«, v osnovnem dokumentu je vidna ikona datoteke), uporabimo makroje ali formule pri oblikovanju vsebine dokumenta, določen del podatkov pa prervorbi v format PDF/A postane nedostopen (glej poglavje 6.6. Prervorba besedilnega dokumenta iz izvirne v PDF/A obliko zapisa). S tem pa je kršeno tudi načelo dostopnosti pri dolgoročni hrambi gradiva v digitalni obliki, ki določa: »Gradivo oziroma reprodukcija njegove vsebine mora biti ves čas trajanja hrambe zavarovana pred izgubo ali okrnitvijo celovitosti ter dostopna tistim, ki imajo za to ustrezna pooblastila«.
2	Je široko priznan in uveljavljen oziroma uporabljen, ter njegova uporaba pa podpira z na trgu uveljavljeno strojno in programsko opremo.	DA	PDF/A je glede na našo raziskavo splošno priznan in uveljavljen oz. uporabljen format, uporaba pa je podprta s strojno in programsko opremo, uveljavljeno na trgu (glej poglavje 6.3. Pregled razširjenosti PDF/A v mednarodnih institucijah).
3	Neposredno uporaben za reprodukcijo vsebine ali enostavno prervorjen v obliko, ki je neposredno uporabna.	DA NE	Ne glede na to, da zakon to lastnost navaja kot eno, pa smo jo razbili na dva dela, saj se nanaša na različni lastnosti: neposredno uporaben za reprodukcijo vsebine enostavno prervorjen v obliko, ki je neposredno uporabna. Pri razumevanju izpolnjevanja te lastnosti smo si pomagali še s slovarjem slovenskega knjižnega jezika ⁸⁴ našli smo to razlago besed: reproducirati: ustvarjati kaj tako, kot že obstaja v stvarnosti; neposreden -dna -a prid., neposrednejši (-e-e): ki je brez posredovanja, brez česa vmesnega: neposreden stik, vpliv; uporaben -bna -o prid., uporabnejši (u a): ki se da koristno uporabiti, porabiti; ki je v takem stanju, da se lahko uporablja; reprodukcija je zelo natančna kopija originala (http://bs.wikipedia.org/wiki/Reprodukcija). Format PDF/A je neposredno uporaben za reprodukcijo vsebine. To lahko dokazujemo tudi s tem, da je že vgrajen v posamezne aplikacije kot npr. Adobe Acrobat, pisarniške zbirke MS Office itd. Drugi del te lastnosti pa PDF/A ne izpolnjuje, saj po prervorbi ni zagotovljena neposredna uporaba podatkov, ki so v vgrajeni darotki (glej primer iz poglavja 6.4. – vgrajena daroteka).
4	Omogoča samodejno prervorbo iz najpogostejše uporabljanih izvirnih oblik zapisa s samodejno zaznavo in poročanjem o nepredvidenih dogodbah oziroma napakah pri prervorbi.	NE	V slovarju slovenskega knjižnega jezika ima beseda samodejen ta pomen: ki opravlja delo sam, brez človekovega sodelovanja; avtomatičen. PDF/A omogoča samodejno prervorbo, vendar je ob ugotovljenih napakah pri prervorbi potrebna človeška intervencija (glej poglavje 6.4. Prervorba besedilnega dokumenta iz izvirne oblike zapisa v PDF/A, primer Bergmann).

⁸³ Ugotovitev pomeni: DA – kriterij je izpolnjen, NE – kriterij ni izpolnjen.

⁸⁴ SAZU in ZRC, Inštitut za slovenski jezik Frana Ramovša in avtorji. Založnik DZS, d.d.. Elektronska izdaja v 1.0. Prikaz s programom ASP32 – pregledovalnik podatkovnih zbirk v 1.40.

5	<i>Je neodvisen od posamezne programske ali strojne opreme oz. ravnja okolja.</i>	DA	Glej poglavje 5.2. Značilnosti PDF/A 1 (http://www.aiim.org/documents/standards/19005-1_FAQ_PDF (najdeno 19. 11. 2011)).
6	<i>Glede na stanje stroke obstaja velika stopnja verjetnosti, da zagotavlja varno hrambo več kot pet let.</i>	NE	ZVDA GA pojem »varna hramba« določa kot hrambo izvirnega gradiva v digitalni obliki z zagotovljenimi načeli dostopnosti, uporabnosti in celovitosti (26. člen). Na podlagi primerov iz poglavja 5.4. Pretvorba besedilnega dokumenta iz izvirne oblike zapisa v PDF/A lahko zaključimo, da za zdaj PDF/A teh načel v določenih primerih ne zagotavlja še ob sami pretvorbi, zato je ne bo mogel zagotavljati tudi v naslednjih letih. Po dostopnih virih sodeč pa lahko sklepamo, da je to problem, ki bo z naslednjimi različicami PDF/A odpravljen. Vendar ta trditve velja samo za določene primere (primer vgrajenih datotek v datoteko, formule, makroji), kjer se je sicer ohranila struktura (videz, oblika) dokumenta, njen kontekst pa ne v celoti (glej International Council on Archives, 2005). Dodaten problem predstavlja tudi dejstvo, da je PDF/A v lasti podjetja Adobe Systems Incorporated, ki nadejra razvoj oblike zapisa, ni pa obvezano, da dolgoročno obliko zapisa ohranja in razvija ali celo objavlja specifikacijo prihodnjih različic. Poleg tega lahko oblika zapisa vsebuje elemente, ki so nekompatibilni s prejšnjimi različicami oblike zapisa. Težava je v tem, da je specifikacija oblike zapisa PDF javna, to pa omogoča različnim razvijalcem programske opreme izdelavo programov, ki izdelajo dokument v obliki zapisa PDF (med drugim tudi najnovejši paket Microsoft Office 2007/2010 vključuje možnost pretvorbe v obliko zapisa PDF). Ti razvijalci lahko kljub specifikaciji oblikujejo zapis PDF na različne načine, to pa resnič povzroča manjše nekompatibilnosti določenih dokumentov, zapisanih v obliki zapisa PDF. Razvijalci lahko vključijo elemente, ki jih Adobe ni predvidel (glejte tudi poglavje 6.1. Zgodovinski razvoj formata PDF/A).
7	<i>Omogoča po današnjih strokovnih predvidevanjih po tem obdobju pretvorbo v novo, takrat določeno obliko zapisa za dolgoročno hrambo.</i>	DA	Če verjamemo napovedim podjetja Adobe in združenja AIIM (http://www.aiim.org/documents/standards/19005-1_FAQ_PDF), lahko z gotovostjo potrdimo, da PDF/A ta kriterij izpolnjuje. Kljub temu pa moramo biti previdni, saj je PDF/A ne nazadnje lastniška oblika formata. Če upoštevamo še druga dejstva (glej utemeljitev 6, drugi odstavek), pa težko pritrjujemo, da PDF/A ustreza tej lastnosti.
8	<i>Temelji na mednarodnem, državnem ali splošno priznanem in praviloma odprtem standardu, če obstaja.</i>	DA	PDF/A je ISO standard (glej poglavje 6.1. Zgodovinski razvoj formata PDF/A).

Tabela 4: Pregled izpolnjevanja splošnih lastnosti formata PDF/A-1, kot jih določajo slovenski predpisi

Dosedanji izsledki nas torej vodijo k temu, da strategija prevorbe gradiva iz ene digitalne oblike v drugo ni dovolj dobra strategija za dolgoročno hranjenje.

Glede na zmogljivost trenutnih formator zapisa, ki jih je stroka opredelila kot primerne za dolgoročno hrambo, in dejstvo, da so še vedno v razvoju, za zdaj predlagamo, tudi obranjanje izvirnih oblik dokumentov poleg pretvorjenih.

Verjetno gre za prehodni predlog za čas, dokler obstoječi formati, predvideni za dolgoročno hrambo, ne bodo izpopolnjeni. Npr. Adobe zagotavlja, da bo v prihodnjem letu objavil različico PDF/A-3,

ki bo odpravljala težave, zaznane pri našem primeru pretvorbe.⁸⁵ Zato predlagamo, da organizacije za zdaj redno spremljajo in pregledujejo razvoj formatov, da bodo zagotovile izbiro najboljšega za svoje potrebe.

⁸⁵ Dne 27. 10. 2011 je avtorica tega prispevka dobila v svoj e-poštni predal odgovor g. Thomasa Zellmanna, PDF Association (e-mail: thomas.zellmann@pdfa.org) s to vsebino: »PDF/A-3 will support any embedded file like your Word and Excel example. This was an requirement from ARCHIVES to the ISO committee but PDF/A-3 is planned for 2012 and not published yet«.

V prid temu predlogu govori tudi izkušnja nizozemske nacionalne knjižnice pri projektu DARE (Digital Academic Repositories), ki ga v svojem poročilu navajata Judith Rog in Caroline van Wijk⁸⁶. V okviru projekta so zbirali publikacije iz različnih skladišč shranjenih publikacij (angl. *repositories*), navajata primer nizozemske univerze. Večina zbranega gradiva je bilo v formatu PDF, manjši del člankov pa formatih MS Word, MS PowerPoint in WordPerfect. Odločili so se, da bodo te dokumente pretvorili v format PDF/A. Z metodologijo ocenjevanja primernosti formatov za dolgoročno hrambo na podlagi vnaprej določenih kriterijev so ugotovili, da dokumenti v Wordu izpolnjujejo te kriterije 22, PDF/A pa 89-odstotno. Glavna razlika med njima pa je bila v odprtosti (angl. *openness*), privzetosti (angl. *adoption*) in odvisnosti (angl. *dependencies*). Po teh lastnostih Word ni imel bistveno nižjega rezultata kot PDF/A. Glede na ta rezultat so se odločili, da bodo za dolgoročno hrambo ohranili tako izvirno obliko dokumentov kot tudi pretvorjeno obliko.

Nadalje Fanning⁸⁷ v poročilu o uporabi PDF in PDF/A navaja, da samo uvedba oz. uporaba teh formatov ne bo zagotavljala dolgoročne hrambe digitalnih dokumentov.

Fanning:

Organizacija bo zagotovila hranjenje svojih digitalnih dokumentov le, če bo poleg implementacije in uporabe PDF/A formata sprejela tudi menedžerski program za upravljanje dokumentov ter formalno uvedla politiko in procese upravljanja dokumentov.

Organizacijam predlaga, naj si pri sprejetju menedžerskega programa za upravljanje dokumentov ter formalni uvedbi politik in procesov upravljanja dokumentov pomagajo z različnimi standardi, kot so ISO 15489-1, ISO 15489-2 ter ISO 14721. Z uporabo teh standardov pri uvedbi programa za upravljanje dokumentov bo uporaba formata PDF/A učinkovitejša in bo lahko zagotavljala dolgoročno hrambo digitalnih dokumentov.

Slovenija je šla korak dalje in je, kot predlaga B. A. Fanning, to udeležila s sprejetjem ZVDAGA. Ta v smislu proaktivnosti⁸⁸ **vpeljuje pripravo in sprejetje notranjih pravil za zajem in digitalno hrambo**, za javnopravne osebe pa tudi zahteva, da

jih potrdijo pri Arhivu RS.⁸⁹ Po definiciji⁹⁰ zakona so »notranja pravila za zajem in hrambo gradiva v digitalni obliki« pravila, ki jih kot svoj interni pravni akt sprejme organizacija v zvezi s hrambo svojega gradiva. Njihova osnovna naloga je opisati vse delovne procese, ki jih izvajajo zaposleni v organizaciji za zagotovitev ustreznega in učinkovitega zajema in hrambe gradiva v digitalni obliki. V tem primeru bi se lahko izognili izgubi dela vsebine pretvorjenih dokumentov v novo digitalno obliko, saj bi lahko s pravili določili ravnanja ob uporabi vgrajenih dokumentov, makrojev in formul.

Slovenska javna arhivska služba (SJAS) je pripravi, vlada RS pa v letu 2010 tudi potrdila strategijo slovenskega javnega elektronskega arhiva e-ARH.si.⁹¹ V njej je SJAS sicer zapisala, da bo uresničevala aktivno politiko sodelovanja z ustvarjalci, da bo pravočasno in učinkovito poskrbljeno za pripravo arhivskega gradiva v digitalni obliki. Na ta način želi zmanjšati tveganja, da v času, ko je gradivo v digitalni obliki še pri ustvarjalcu, ne bodo zagotovljena načela varne hrambe tega gradiva in bi bilo v praksi kot posledica to gradivo izgubljeno. V nasprotju z našim predlogom, da bi ob pretvorbi ohranili tudi izvirne oblike dokumentov, je strategija usmeritev SJAS po tej strategiji, da se arhivsko gradivo v digitalni obliki od ustvarjalcev ne bo prevzemalo v izvornem formatu (običajni oz. produkcijski obliki), razen če to ni že oblika, ki je primerna za dolgoročno hrambo.⁹² Je pa v strategiji zapisano, da bo oblika zapisa (format) za prevzemanje glede na posamezen tip gradiva, predpisala v konkretnih navodilih za odbiranje, ki bodo temeljila na enornih tehnoloških zahtevah (ETZ).

6. Viri in literatura

Viri

About the Global Digital Format Registry. Dostopno na: <http://www.gdfr.info/about.html> (7.11. 2011).

Bearman, David: Capturing Records' Metadata: Unresolved Questions and Proposals for Research. V: *Archives and Museum Informatics* (1997), vol. 11/3–4, str. 271–276.

⁸⁶ Prim. *Evaluating File Formats for Long-term Preservation.*

⁸⁷ Prim. Fanning: *Technology Watch Report.*

⁸⁸ Načelo proaktivnosti razumemo kot zgodnje zaznavanje oz. odkrivanje nevarnosti in pravočasno reakcijo za odpravo teh nevarnosti.

⁸⁹ ZVDAGA, 18. in 19. člen.

⁹⁰ Prav tam, 2. člen.

⁹¹ *Strategija razvoja slovenskega javnega elektronskega arhiva e-ARH.si*, 32 strani.

⁹² Prav tam, str. 17.

Bearman, David: *Reality and Chimeras in the Preservation of Electronic Records*. V: *D-Lib Magazine* (1999), vol. 5 (4). Dostopno na:

Digital preservation. Dostopno na: http://en.wikipedia.org/wiki/Digital_preservation#cite_note-1#cite_note-1 (29.10.2011).

Drümmer, Olaf: *What is PDF/A-2?* Dostopno na: http://www.drueemmer.com/_wp/?p=27 (18.11.2011).

Frappaolo, Carl: *AIIM Market IQ study. Content Creation and Delivery: The On-Ramps and Off-Ramps of ECM*. Dostopno na:

<http://www.aiim.org/pdfdocuments/35498.pdf> (18.11.2011).

How to answer questions in PDF documents? Dostopno na:

<http://www.a-pdf.com/faq/how-to-answer-questions-in-pdf-documents.htm> (16.11.2011).

<http://www.dlib.org/dlib/april99/bearman/04bearman.html> (17.10.2011).

<http://www.nationalarchives.gov.uk/PRONOM/Default.aspx> (7.11.2011).

<http://www.businessdictionary.com/definition/proprietary-standard.html> (26.10.2011).

Leonhardt, David: John Tukey, 85, Statistician; Coined the Word 'Software'. V: *New York Times*, Obituaries, July 28, 2000. Dostopno na:

<http://www.nytimes.com/2000/07/28/us/john-tukey-85-statistician-coined-the-word-software.html> (10. 2011).

Making the case for PDF/A and Adobe® Acrobat®. Dostopno na:

<http://www.adobe.com/enterprise/pdfs/pdfaforAcrobat.pdf> (18.11.2011).

Merwe, Adèle, & Deventer, Martie: *Planning an effective digital preservation strategy for a research organization*. Dostopno na:

http://www.ais.up.ac.za/digi/docs/avdmerwe_paper.pdf (26.10.2011).

Methodologies for the Long-Term Preservation of and Access to Software-Dependent Electronic Records. Dostopno na:

http://www.sdsc.edu/NARA/Publications/nhprc_latest.pdf (2.11.2011).

Mind the Gap: Digital Preservation Needs in the UK. Dostopno na:

<http://www.dpconline.org/advocacy/mind-the-gap> (17.10.2011).

Nearly all archiving projects use PDF/A. Dostopno na:

<http://www.computeruser.com/pressreleases/nearly-all-archiving-projects-use-pdf/a.html> (18.11.2011).

PDF/A. Dostopno na:

<http://en.wikipedia.org/wiki/PDF/A> (15.11.2011).

Proprietary standard. Dostopno na:

Rog, Judith, & Wijk, Caroline. *Evaluating File Formats for Long-term Preservation*. The Hague: National Library of the Netherlands, 2012. Dostopno na:

http://www.kb.nl/hrd/dd/dd_links_en_publicaties/publicaties/KB_file_format_evaluation_method_27022008.pdf (18.1.2012).

Rothenberg, Jeff: *Avoiding technological quicksand: Finding a viable technical foundation for digital preservation*. A report to the council on library and information resources (1998). Dostopno na:

<http://www.clir.ORG/pubs/reports/rothenberg/contents.html> (4.11.2011).

Sporočilo Komisije Evropskemu Parlamentu, Svetu, Evropskemu Ekonomsko-Socialnemu Odboru in Odboru Regij - iz 2010: digitalne knjižnice {SEC(2005) 1194} {SEC(2005) 1195} /* KOM/2005/0465 končno */. Dostopno na:

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2005:0465:FIN:SL:IT:HTML> (20.10.2011).

The InterPARES 2 Project Dictionary. Dostopno na:

http://www.interpares.org/ip2/display_file.cfm?doc=ip2_dictionary.pdf (6.11.2011).

The technical registry PRONOM. Dostopno na:

TrID - File Identifier. Dostopno na:

<http://mark0.net/soft-trid-c.html> (7.11.2011).

What is a File Extension? Dostopno na:

<http://filext.com/> (7.11.2011).

WordPerfect. Dostopno na:

<http://en.wikipedia.org/wiki/WordPerfect> (21.6.2011).

Literatura

Arhiviranje, hramba in upravljanje dokumentov: Navodila in praktični primeri pravilne hrambe vseh vrst dokumentarnega gradiva za podjetja in ustanove. (ur. Tatjana Hajtnik). Maribor: Prešernova Založba, Forum Media, 2009–2011, mapa, nevezani listi.

Bergmann, Andreas: *PDF/A in Siemens PLM Environment. 3rd International PDF/A Conference: PDF/A up to date, Long-Term Archiving with PDF*. Berlin: Siemens AG A&D, Germani, 2009, str. 93–96.

Bikson, K. Tora & Frinking, Eric: *Preserving the Present: Toward Viable Electronic Records*. Den Haag: Sdu Publishers. (v delu, ki je dostopen kot RAND Reprint RP-257), 1993.

Conway, Paul: Archival Preservation Practice in a Nationwide Context. V: *The American Archivist* 53 (1990), str. 204–222.

Drümmer, Olaf & Oettler, Alexandra & Seggern, Dietrich: *PDF/A in a Nutshell: Long Term Archiving with PDF*. Berlin: Competence center PDF/A, 2007.

Duranti, Luciana: Concepts, Principles, and Methods for the Management of Electronic Records. V: *The Information Society* 17 (2001), str. 271–279.

Duranti, Luciana: *Diplomatics: New Uses for an Old Science*. Scarecrow Press, 1998.

Enotne tehnološke zahteve ver 2.0, I. del: Uvodna poglavja. Ljubljana: Arhiv Republike Slovenije, 2011.

Enotne tehnološke zahteve ver 2.0, II. del: Enotne tehnološke zahteve za zajem in hrambo gradiva v digitalni obliki. Ljubljana: Arhiv Republike Slovenije, 2011.

Enotne tehnološke zahteve ver 2.0, I–III. Ljubljana: Arhiv Republike Slovenije, 2011.

Fanning, A. Betsy: *Technology Watch Report. Preserving the Data Explosion: Using PDF*. ALIM, 2008.

Folk, Mike & Barkstrom, R. Bruce: *Attributes of File Formats for Long-Term Preservation of Scientific and Engineering Data in Digital Libraries*. ACM, 2003.

Guide for managing electronic records from an Archival Perspective. ICA Studies/Études CIA, International Council on Archives, 1997.

Hedstrom, Margaret & Montgomery, Sheon: *Digital Preservation Needs and Requirements in RLG Member Institutions*. A study commissioned by the Research Libraries Group, 1998.

Hedstrom, Margaret: Understanding Electronic Incunabula: A Framework for Research on Electronic Records. V: *American Archivist* 54 (1991), št. 3, str. 334–354.

ISO 14721:2003: *Space data and information transfer systems - Open archival information system - Reference model*. International Organization for Standardization, 2003.

ISO 15489-1: *Information and Document Management – Records Management (Part 1: General)*. International Organization for Standardization, 2001.

ISO 15489-2: *Information and Document Management – Records Management (Part 1: Guidelines)*. International Organization for Standardization, 2001.

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

PREMIS: *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group*. Preservation Metadata: Implementation Strategies (PREMIS), United States: On-Line Computer Library Center and Research Libraries Group, 2005.

Strategija razvoja slovenskega javnega elektronskega arhiva e-ARH.si. Ljubljana: Arhiv Republike Slovenije, 2010.

Thomassen, Theo: A First Introduction to Archival Science. V: *Archival Science* (2001), vol. 1, str. 373–385.

Uredba o varstvu dokumentarnega in arhivskega gradiva /UI DAG-A/. Uradni list RS, št. 86/2006.

Zakon o arhivskem gradivu in arhivih /ZAGA/. Uradni list RS, št. 20/97, 32/97 in 24/2003.

Zakon o varstvu dokumentarnega in arhivskega gradiva ter arhivih /ZV DAG-A/. Uradni list RS, št. 30/2006.

Zusammenfassung

TAUGLICKEITSANALYSE DES PDF/A-FORMATS ZUR LANGFRISTIGEN AUFBEWAHRUNG VON ARCHIVMATERIAL IN DIGITALER FORM

Angeichts der immer größeren Menge an Dokumenten im digitalen Originalformat setzen sich die Fachleute weltweit mit der Frage der langfristigen Aufbewahrung auseinander. Es ist nämlich eine Tatsache, dass sich die originalen Aufzeichnungsformate nicht zur langfristigen Aufbewahrung eignen, da sie von der Software, die sich unaufhörlich ändert oder sogar verschwindet, allzu sehr abhängig sind. Die derzeit am weitesten verbreitete Strategie der langfristigen Aufbewahrung digitaler Dokumente ist die Konvertierung vom Originalformat in ein zur dauerhaften Aufbewahrung geeignetes Aufzeichnungsformat. Auch die slowenischen Bestimmungen auf dem Gebiet der Verwaltung, Erfassung und Aufbewahrung von Schrift- und Archivgut schreiben eine Konvertierungsstrategie vor, die eine langfristige Aufbewahrung der digitalen Dokumente ermöglicht. Diese definieren die bei der langfristigen digitalen Aufbewahrung zu berücksichtigenden Grundsätze und auch die Kriterien, die die zur langfristigen Aufbewahrung geeigneteren Formate erfüllen müssen.

Da die Konvertierung von einem digitalen Format in ein anderes weltweit und auch in Slowenien die derzeit am meisten verfolgte Strategie der langfristigen Aufbewahrung digitaler Informationen ist,