

1.08 Objavljen znanstveni prispevek na konferenci  
1.08 Published Scientific Conference Contribution

Tatjana HAJTNIK\*

## **PRETVORBA E-DOKUMENTOV V FORMATE ZA DOLGOROČNO HRAMBO NA OSNOVI STANDARDA ISO 13008**

### **Izvleček:**

Ob vse večjih količinah e-zapisov njihovo dolgoročno ohranjanje ni zgolj naša želja, je potreba, saj bodo ti zapisi nujno potrebni za kasnejše dokazovanje pravnih, finančnih in podobnih stanj, dokazovanje sprejemanja in izvajanja poslovnih odločitev ali ohranjanje kulturne dediščine. Vendar pa dolgoročno ohranjanje e-zapisov ni primerljivo z ohranjanjem zapisov na papirju, saj sta njihov dostop in uporaba v prvi vrsti odvisna od tehnologije in njenega hitrega razvoja. Avtorica v prispevku predstavi strategijo pretvorbe e-dokumentov kot trenutno v svetu najbolj razširjeno strategijo za dolgoročno ohranjanje e-dokumentov. Teoretični okvir procesa pretvorbe predstavlja leta 2012 sprejet standard ISO 13008, ki ga avtorica v posameznih delih specializira in podpre s konkretnimi rešitvami ter s tem oblikuje celovit pristop k pretvorbi e-dokumentov v formate, ki so primerni za dolgoročno hrambo.

### **Ključne besede:**

e- dokumenti, strategija pretvorbe, format, dolgoročna elektronska hramba, ISO 13008

### **Abstract:**

#### **Converting Electronic Documents into Formats for Long-term Preservation on the Basis of Standard ISO 13008**

Considering the ever-increasing number of digital records that are being produced daily, the issue of their long-term preservation is of paramount importance, since such digital records will in the future be needed to prove our legal and financial conditions, to explain our processes of making and implementing business decisions, and to preserve our cultural heritage. There is, however, no comparison between the task of preserving paper records and that of long-term preservation of digital records, since the access and usage of the latter depends primarily on the technology and its rapid development. In the article, the author presents the strategy of converting digital documents as currently the most widespread strategy for long-term preservation of digital records. Theoretical framework for such conversion is the standard ISO 13008, adopted in 2012, which in places is fully specified by the author and supported by concrete solutions. The author thus develops an holistic approach to converting digital records into formats that are suitable for long-term preservation.

### **Keywords:**

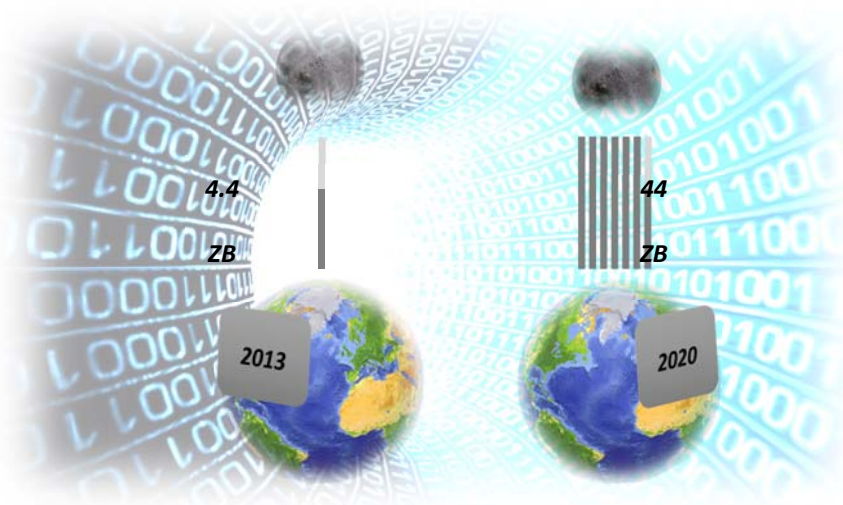
digital records, conversion strategy, format, long term digital preservation, ISO 13008

---

\* Dr. Tatjana Hajtnik, univ. dipl. ing. rač. in inf., Arhiv Republike Slovenije, Zvezdarska 1, 1000 Ljubljana, Slovenija, kontakt: [tatjana.hajtnik@gov.si](mailto:tatjana.hajtnik@gov.si)

## 1. UVOD

V zadnjih letih se način dela in komunikacij spreminja, delo z elektronskimi zapisi in podatkovnimi bazami izpodriva tradicionalno pisarniško poslovanje, elektronska pošta nadomešča klasično pošiljanje papirnatih dopisov, papirnate obrazce in vloge nadomeščajo elektronske vloge in še bi lahko naštevali. Znana mednarodna analitska hiša International Data Corporation (2014) je na podlagi izvedene raziskave leta 2014 napovedala, da se bo količina digitalnih zapisov od leta 2013 (4,4 bilijona GB) do 2020 (44 bilijonov GB) vsaki dve leti več kot podvojila. Za lažjo predstavo: če bi digitalne zapise predstavili v obliki skladovnice, bi ta leta 2013 predstavljala dve tretjini poti od Zemlje do Lune, leta 2020 pa kar 6,6 poti (Slika 1).



**Slika 1: Digitalni svet je ogromen in z leti eksponentno narašča**

Ne glede na novo, sodobno elektronsko obliko ti zapisi predstavljajo ključni vir za poznavanje in obvladovanje gospodarstva, dragoceni so za različne raziskave in izobraževanja, zgodovino, znanost in humanistiko, ustvarjalne in kulturne dejavnosti ter javni red. Zato moramo, tako kot stoletja zaradi različnih potreb ohranjamo zapise v fizični (papirni) obliki, za sedanje in prihodnje generacije ohranjati tudi zapise v elektronski obliki ter zagotavljati dostop do njih kot tudi njihovo uporabnost in verodostojnost. V nadaljevanju se bomo omejili in kot temeljno enoto elektronskega zapisa obravnavali dokument, ki je ustvarjen in shranjen z elektronskimi sredstvi v določenem formatu.

Ohranjanje e-dokumentov za daljše časovno obdobje na način, da ohranimo njihovo enakost *izvirnikom*<sup>1</sup> in s tem tudi njihovo verodostojnost, že vrsto let predstavlja enega večjih izzivov, saj ni enostavno primerljivo z dolgoročno hrambo dokumentov v fizični (papirni) obliki. Prikaz in interpretacija slednjih namreč nista odvisna od strojne ali programske opreme, saj je znano, da se lahko dokumenti na kvalitetnem arhivskem papirju ali na mikrofilmu ohranjajo več kot 500 let (Puglia, 1999). Pri dolgoročni hrambi

<sup>1</sup> **Izvirnik:** »2. prvi, prvotni izvod, primerek listine: izgubiti izvirnik; primerjati prepis z izvirnikom; fotografija izvirnika« (SSKJ, 2000).

e-dokumentov pa moramo upoštevati več dejavnikov. Na prvo mesto prav gotovo sodi odvisnost e-dokumentov od opreme, tako strojne kot programske, ki jo potrebujemo za njihovo predstavitev v človeku razumljivi obliki, ter po drugi strani hitrost spreminjanja te opreme. Razvoj strojne opreme ima za posledico tudi hitro spreminjanje nosilcev e-dokumentov, razvoj programske opreme pa hitro spreminjanje oblike zapisov. Medtem ko pri dokumentih v fizični obliki neukrepanje v primeru njihove dolgoročne hrambe običajno ni škodljivo, pa hiter tehnološki razvoj nedvomno zahteva takojšnje ukrepanje in proaktivno ravnanje pri ohranjanju e-dokumentov. Tako je Rothenberg (1999) že pred leti ugotovil, da je težave pri ohranjanju e-dokumentov možno reševati z različnimi strategijami<sup>2</sup>, kot so npr. emulacija (angl. *emulation*), osveževanje medijev (angl. *refreshing bits media migration*), tiskanje na papir (angl. *printing*), ohranjanje tehnologije (angl. *technology preservation*) in snemanje (angl. *filming*).

V današnjem času pa se vedno bolj uveljavlja strategija pretvorbe e-dokumentov iz ene v drugo obliko zapisa<sup>3</sup> oz. formate, ti pa naj bi bili primerni za dolgoročno hrambo. Strategija pretvorbe je opredeljena v različnih virih (npr. Waters & Garrett (1996), ISO 15489-1, ISO 30300, ISO 13008), vsem pa je skupno, da gre pri pretvorbi za proces spreminjanja e-dokumentov in hkrati ohranjanja njihovih lastnosti (z malo ali brez izgube). V Sloveniji je, kadar je treba e-dokumente hraniti več kot pet let, ta strategija kot obvezna določena celo s predpisi (ZVDAGA, 11. člen). Na strategijo pretvorbe e-dokumentov v formate, ki so primerni za dolgoročno hrambo, gledamo torej kot na enega od ukrepov, s pomočjo katerega lahko dolgoročno ohranjamo e-dokumente, zmanjšujemo tveganja, ki izhajajo iz tehnološke ali ekonomske<sup>4</sup> zastarelosti, omogočimo ponovno uporabo e-dokumentov in povečamo njihovo interoperabilnost.

Vendar pa ne smemo zanemariti dejstva, da se s hitrim tehnološkim razvojem spreminjajo tudi formati, čeprav so v določenem obdobju sprejeti kot primerni za dolgoročno hrambo e-zapisov. Tipičen primer, ki to potrjuje, je razvoj formata PDF oz. kasneje na njegovi osnovi temelječega PDF/A, ki je v svetu že vrsto let med bolj uveljavljenimi za dolgoročno hrambo e-dokumentov. Razvoj tega formata traja že vse od leta 1993 in do danes je bila izdana že vrsta njegovih različic. Vsaka od njih ima, vključno z danes zadnjo različico PDF/A-3, poleg prednosti tudi vrsto pomanjkljivosti (Hajtnik, 2012). Posledično lahko to pomeni, da se pri pretvorbi e-dokumenta v ta format lahko del podatkov izvirnega e-dokumenta izgubi, s tem pa se lahko bistveno zmanjša tudi njegova verodostojnost. Vse omenjeno kaže na to, da ni dovolj, da e-dokumente pretvarjamo v formate, ki so primerni za dolgoročno hrambo, temveč mora biti celoten

<sup>2</sup> **Strategija:** »...postopki, načini za doseg kakega cilja« (SSKJ, 2000).

<sup>3</sup> Slovenski predpisi format poimenujejo »oblika zapisa« in jo opredeljujejo »kot organizacijske in tehnološke značilnosti zapisa, ki določajo, kako je vsebina zapisana, hranjena in prikazana v procesu hrambe« (ZVDAGA, 2. člen). Format je odvisen od vrste dokumenta, njegove vsebine in načina uporabe. Določeni programski opremi omogoča interpretirati podatke v dokumentu; je način predstavitve teh podatkov. ZVDAGA (11. in 12. člen) za obliko zapisa ob nastanku oz. v času ustvarjanja gradiva (dokumentov) uporablja izraz »običajna digitalna oblika«, z Enotnimi tehnološkimi zahtevami (2013) pa se ta oblika zapisa imenuje produkcijska oblika zapisa.

<sup>4</sup> Tehnološka zastarelost nastane, ko tehnologija za branje ali dostop do informacij ni več na voljo. Ekonomska zastarelost pa nastane, ko postanejo stroški vzdrževanja (ali ponovnega pridobivanja) tehnologije za dostop do informacij v nekem formatu preveliki.

postopek pretvorbe skrbno načrtovan. Ključne dejavnike, ki vplivajo na uspešno izvedbo pretvorbe, predstavljajo določitev načel, s katerimi dolgoročno hrambo e-dokumentov lahko opredelimo kot varno, ter izbire ustreznih formatov in programskih orodij za izvedbo pretvorbe. V prispevku je prikazan celovit pristop k izvajanju postopka pretvorbe e-dokumentov v formate, primerne za dolgoročno hrambo, ki temelji na *standardu ISO 13008* in je v posameznih fazah specializiran in konkretiziran s posameznimi rešitvami. S predstavljenim celovitim pristopom k pretvorbi e-dokumentov lahko povečamo možnosti njihove *varne dolgoročne hrambe* ter ponovne uporabe.

## 2. VARNA DOLGOROČNA HRAMBA E-DOKUMENTOV

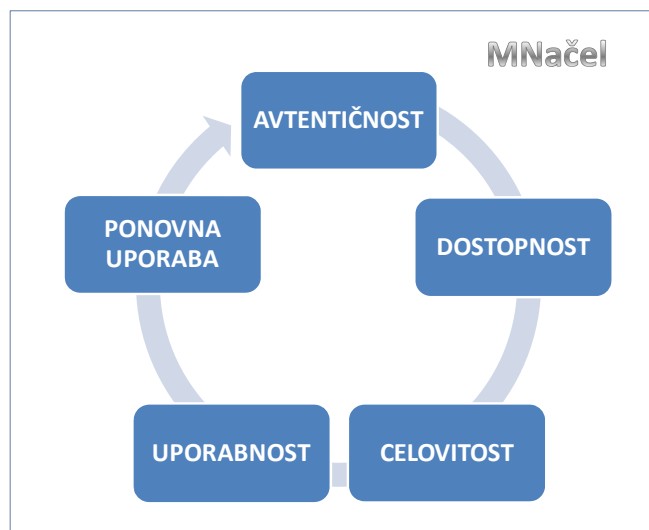
Če povzamemo različne definicije (npr. ALCTS, 2007; ISO 14721), potem je cilj dolgoročne hrambe e-dokumentov njihovo ohranjanje tako, da bodo ves čas hrambe, ne glede na spreminjanje tehnologij, dostopni in verodostojni, kar pomeni, da morajo ohranjati *bistvene lastnosti izvirkov*. Če izhajamo iz tega, da so bistvene lastnosti dokumentov tiste, ki so bistvene za njihov pomen, uporabo ali organizacijsko vrednost, potem jih je treba ohraniti tudi pri kakršnikoli posegih v e-dokumente, njihovem prenašanju v druga tehnološka okolja ali pretvarjanju v druge formate. Nekaj osnovnih virov (ICA, 1997; Rothenberg in Bikson, 1999; Thomassen, 2001; ISO15489, 2001; Duranti, 1998, 2001a in 2001b; Bearman, 1999; Cox, 2001) kot bistvene lastnosti e-dokumentov izpostavlja njihovo *vsebino* (angl. *content*), *kontekst* (angl. *context*), *strukturo* (angl. *structure*) ter *videz in obnašanje* (angl. *look and feel*). Opisujemo jih z metapodatki<sup>5</sup>, ki jih k e-dokumentom dodajamo kot dodatne informacije. Za potrebe opisa bistvenih lastnosti e-dokumentov obstaja več različnih metapodatkovnih shem, strategij in standardov, npr. *METS (Metadata Encoding & Transmission Standard)*, *Dublin Core (ISO 15836)* in *PREMIS Data Dictionary for Preservation Metadata*.

Katere od lastnosti e-dokumentov, opisane z metapodatki, so pomembnejše oz. sodijo med bistvene, je odvisno tudi od samega tipa e-dokumentov (npr. e-pošte, tekstovnih dokumentov, preglednic in podatkovnih baz), v prvi vrsti pa so določene s posameznim poslovnim procesom. Vsaka organizacija mora sama identificirati in opredeliti, katere lastnosti e-dokumentov so bistvene, in se osredotočiti na njihovo ohranjanje, kar je ključno za *varno dolgoročno hrambo e-dokumentov*.

V nadaljevanju bomo dolgoročno hrambo e-dokumentov opredelili kot varno, če bodo dosežena načela, kot so opredeljena v modelu **MNačel** (Hajtnik, 2016). Ta predstavlja sistematično analizirana in sintetizirana načela, ki jih izpostavljajo posamezni strokovnjaki, mednarodni projekti, predpisi in standardi. Kot ključna načela varne dolgoročne hrambe e-dokumentov model opredeljuje njihovo *dostopnost*, *uporabnost*, *celovitost*, *avtentičnost* in *ponovno uporabo* (Slika 2).

---

<sup>5</sup> Npr. metapodatke, kot potrebne za podporo poslovanju in procesu upravljanja z dokumenti, opredeljuje standard ISO 23081 (2006).



**Slika 2: Model načel za zagotavljanje varne e-hrambe (MNačel) (Vir: Hajtnik, 2016)**

Za doseganje teh načel mora vsaka organizacija sama določiti, katere so tiste bistvene lastnosti izvirnih e-dokumentov, ki jih je treba ob njihovi pretvorbi ohraniti, da bodo tretje stranke zaupale v verodostojnost pretvorjenih dokumentov (Hajtnik, 2016). S tem je povezana tudi odločitev, katere metapodatke se morda ne bo ohranilo, kar izhaja iz ocene njihove potrebe za nadaljnje poslovanje. Pri tem naj velja pravilo, da bolj kot so dokumenti pomembni (večjo imajo vrednost), več njihovih bistvenih lastnosti je treba ohranjati in s tem zagotavljati njihovo nadaljnjo avtentičnost. V primeru odločitve, da se ne ohrani vseh metapodatkov, je treba to skrbno in natančno dokumentirati ter dobiti ustrezno potrditev s strani pooblaščenih oseb. Pri teh odločitvah so lahko v pomoč različni standardi (npr. ISO 23081) in uveljavljene metapodatkovne sheme (npr. METS), na podlagi katerih se lahko izdelata lastne metapodatkovne sheme za posamezne tipe e-dokumentov. Pri odločitvah, katere bistvene lastnosti e-dokumentov je treba ohraniti, izhajamo iz poslovne vrednosti e-dokumentov, zahtev predpisov ali drugih zahtev.

### **3. V KATEREM FORMATU SO ZAPISANI E-DOKUMENTI?**

E-dokument ni nujno v formatu, za katerega mislimo, da je. Osnovno možnost nam sicer ponuja prepoznavanje končnice imena datoteke (npr. končnica .doc pomeni format programa Microsoft Word), vendar ta način ni stoočstotno zanesljiv. Končnico imena datoteke lahko namreč tudi zelo enostavno, namerno ali nenamerno, spremenimo. Recimo, v primeru zahteve za pretvorbo e-dokumentov v format za dolgoročno hrambo bi lahko dokumentom s končnico .docx to enostavno spremenili v .pdf. V tem primeru bi lahko razumeli, da je dokument zapisan v formatu PDF, kar pa ne bi držalo. Nekatera od programskih orodij, kot npr. File-Extensions.org (2015), identificirajo format na podlagi priponke, navedejo programske opreme, ki podpirajo ta format in tudi predlagajo oz. poiščejo pretvornike, ki znajo pretvarjati ta format.

Druga možnost je, da format preverimo s pomočjo posebne programske opreme, ki na podlagi notranje karakteristike (npr. zaporedje bitov, ki se enolično pojavi le v določenih formatih) ugotovi, v katerem formatu je zapisan dokument. Te programske opreme so za identifikacijo formata povezane z za te potrebe posebej ustvarjenimi registri formatov. Slednji predstavljajo zbirke podatkov o formatih, običajno se uporabljajo in posodablajo centralno, uporablja pa jih lahko različna programska oprema. Taki registri omogočajo identifikacijo in validacijo formatov ter ponujajo informacije o možnostih pretvorbe, stopnji tveganja, karakterističnih lastnostih in o načinu prikazovanja.

Med bolj znanimi registri formatov je PRONOM<sup>6</sup>, ki ga je razvil Nacionalni arhiv Velike Britanije. Register je brezplačno spletno dostopen, vsebuje podatke o praktično vseh poznanih formatih na svetu in pripadajoče opise ter o programski opremi, njenih življenjskih ciklih in tehničnih specifikacijah, ki jih podpira (The National Archives, 2011). V Nacionalnem arhivu Velike Britanije so izdelali tudi aplikacijo DROID<sup>7</sup>, ki v povezavi z registrom PRONOM avtomatsko identificira format dokumenta še pred dostopom do dokumenta oz. njegovim odpiranjem. PRONOM za natančen opis ali ob neznanem formatu 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 dolgoročne e-hrambe. Register PRONOM vsebuje tudi seznam aplikacij, s katerimi je mogoče prikazati dokument določenega formata.

#### 4. KATERI FORMATI SO PRIMERNI ZA DOLGOROČNO HRAMBO E-DOKUMENTOV?

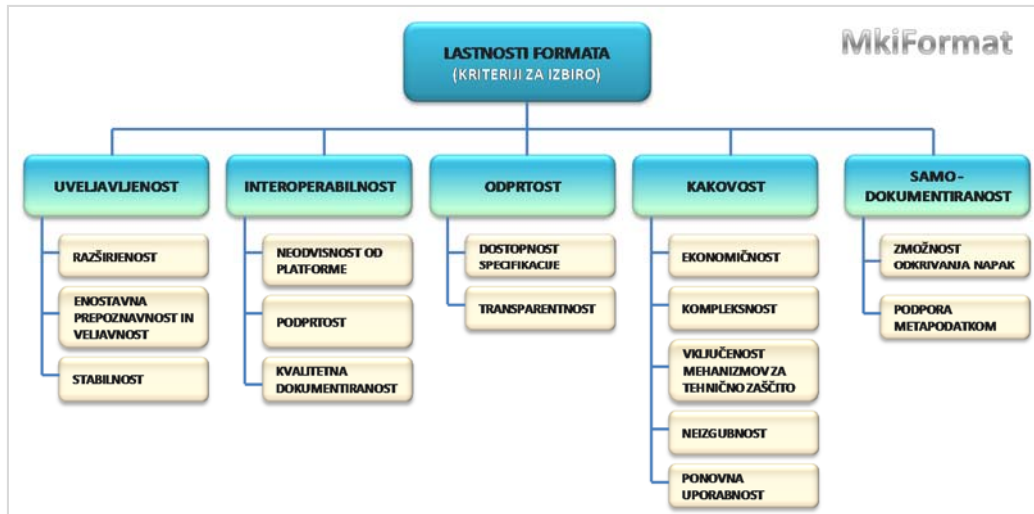
Med več tisoč formati – samo v ZDA so jih našteali več kot 16.000 (Thomas & Martin, 2006) in so danes registrirani v registrih, kot je npr. PRONOM – se seveda postavi vprašanje, kako prepoznati tiste, ki so primerni za dolgoročno hrambo e-dokumentov.

V okviru oblikovanja celovitega pristopa k pretvorbi e-dokumentov v formate, ki so primerni za dolgoročno hrambo, bomo uporabili model kriterijev za izbiro formatov za dolgoročno hrambo *MkiFormat* (Hajtnik, 2016, *poglavje 3.5.4*). Model (Slika 3) je izdelan na podlagi lastnosti formatov, opredeljenih v različnih virih, ki so bile analizirane in sintetizirane ter nato še preverjene in potrjene z empiričnimi podatki, pridobljenimi z anketo v nacionalnih arhivih držav članic EU.

---

<sup>6</sup> Register formatov PRONOM je brezplačno dostopen na spletnem naslovu <http://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=new> (12. 2. 2017).

<sup>7</sup> Aplikacija DROID je brezplačno dostopna na spletnem naslovu <http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>.



Slika 3: Model MkiFormat - kriteriji za izbiro formata (Vir: Hajtnik, 2016)

Model predstavlja pet ključnih lastnosti formatov (*uveljavljenost*, *interoperabilnost*, *odprtost*, *kakovost* in *samodokumentiranost*), te pa predstavljajo širši nabor kriterijev za izbiro formatov. Slednji so na drugem nivoju podrobneje opredeljeni z dodatnimi kriteriji (širši nabor kriterijev). Skladnost formata z modelom MkiFormat in s tem ustreznost za dolgoročno hrambo e-dokumentov je mogoče ugotavljati s kontrolnim seznamom, izdelanim za ta namen (Hajtnik, 2016, *priloga 3*).

## 5. UMESTITEV POSTOPKA PRETVORBE V ŽIVLJENJSKI CIKLUS E-DOKUMENTA

Po Hedstrom in Montgomery (1998) zahteva varna hramba e-dokumentov njihovo upravljanje skozi celoten življenjski cikel. Ta se začne s kreiranjem (ali prejemom) in pridobitvijo statusa izvirnika. Slednje pomeni, da je končna (izvirna) oblika e-dokumenta določena z njegovimi bistvenimi lastnostmi, to je z njegovo vsebino, kontekstom, strukturo, izgledom in obnašanjem, s tem pa je določena tudi njegova vrednost, zaradi katere ga je treba hraniti določeno obdobje. To obdobje je lahko dolgo nekaj desetletij ali več ali pa imajo e-dokumenti celo arhivsko vrednost. Da bodo to obdobje preživeli, je treba za njih ustrezno poskrbeti. Možnost »preživetja« e-dokumentov lahko povečamo s tem, da jih pretvarjamo v formate, primerne za dolgoročno hrambo, to pretvarjanje pa se lahko izvaja v različnih fazah njihovega življenjskega ciklusa. Postopek pretvorbe e-dokumentov, ki imajo status arhivskega gradiva, smo umestili v življenjski cikel e-dokumentov (Slika 4):

- NASTANEK (kreiranje ali prejem e-dokumenta):

Ob nastanku e-dokument dobi status izvirnika, določi se mu rok hrambe:

- rok hrambe < 5 let: dokument shranimo v sistem e-hrambe;
- rok hrambe > 5 let: če je dokument v izvirnem (običajnem) formatu (npr. DOCX), ga *pretvorimo* v format, ki je primeren za dolgoročno hrambo in shranimo v sistem e-hrambe; če je že v formatu, ki je primeren za dolgoročno hrambo, ga shranimo v sistem e-hrambe.

- e-HRAMBA:

V sistem e-hrambe shranimo dokumente, ki jih še potrebujemo pri poslovanju. Rok hrambe e-dokumentov je treba periodično preverjati in v primeru, da:

- *je rok hrambe potekel*: e-dokumente izločimo in uničimo (ohranimo samo evidenco uničenih e-dokumentov);
- rok hrambe *ni potekel* in je *manj kot 5 let od njihovega nastanka*: e-dokumenti ostanejo v sistemu e-hrambe;
- rok hrambe *ni potekel* in je *več kot 5 let od njihovega nastanka*: preverimo, ali je format še primeren za dolgoročno hrambo. V primeru da ni, e-dokumente *pretvorimo* v nov format, ki je primeren za dolgoročno hrambo, nato pa jih prenesemo v arhivski informacijski sistem (AIS), ki je namenjen dolgoročni hrambi e-dokumentov;

- AIS (sistem dolgoročne hrambe e-dokumentov pri ustvarjalcu):

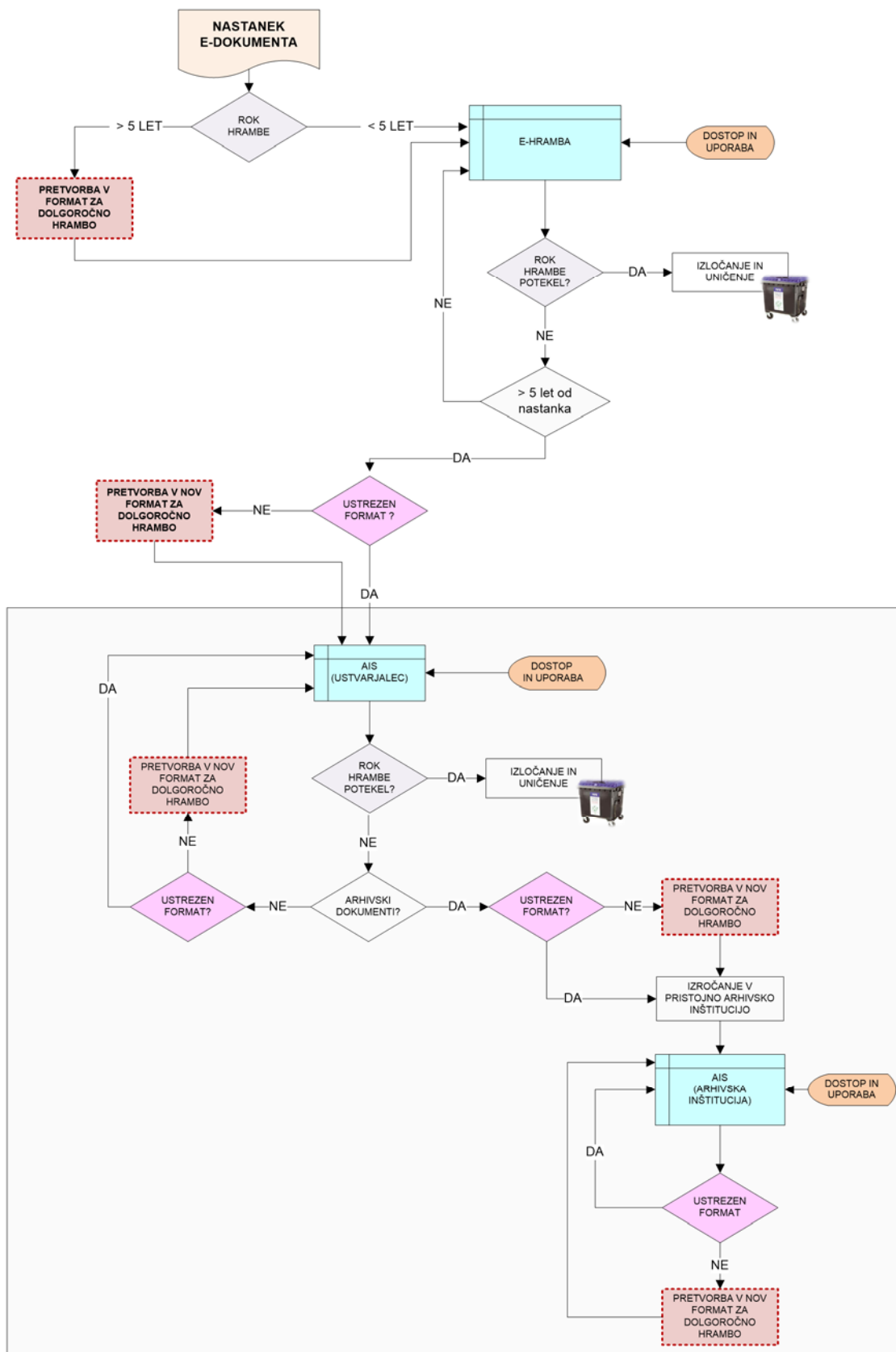
V AIS je treba periodično preverjati rok hrambe e-dokumentov in v primeru, da:

- *je rok hrambe potekel*: e-dokumente izločimo in uničimo (ohranimo samo evidenco uničenih e-dokumentov);
- rok hrambe *ni pretekel*: preverimo ustreznost formata, če ta ni več primeren za dolgoročno hrambo, e-dokumente pretvorimo v nov, ustrezen format.

Postopek pretvorbe se lahko izvaja v fazi *pred izročitvijo* arhivskih dokumentov v pristojno arhivsko inštitucijo ali *takoj po izročitvi*, kasneje pa tudi po potrebi v njenem AIS.

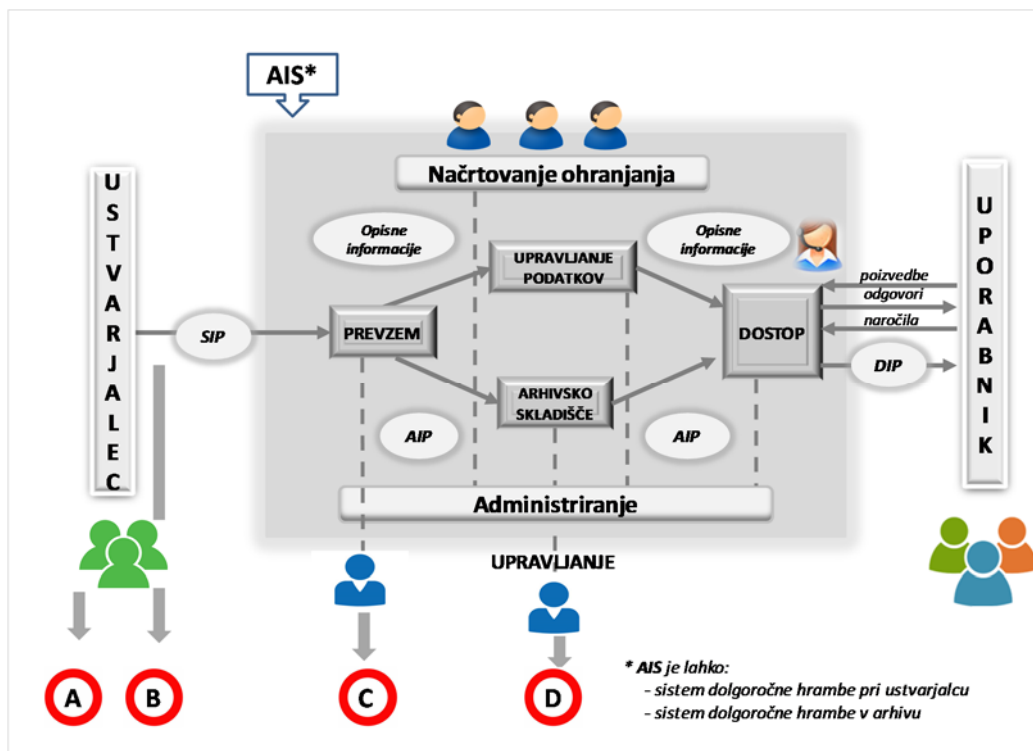
Pri obravnavanju arhivskih e-dokumentov se v praksi z vidika izvajanja pretvorbe e-dokumentov v druge formate soočamo z različnimi situacijami. Običajno so predmet prevzema v arhivske inštitucije e-dokumenti, ki so nastali precej nazaj v preteklosti, v skladu s slovenskimi predpisi je to lahko tudi 30 let od nastanka teh dokumentov (ZVDAGA, 40. člen). To pa lahko tudi pomeni, da se k pretvorbi pristopi šele v fazi izročitve arhivskega gradiva v pristojne arhivske inštitucije. Ti dokumenti so lahko še vedno v formatih, ki pa so že zastareli in z obstoječimi tehnologijami nezdružljivi. V tem primeru se pretvorba lahko izvaja pred izročitvijo (izvaja jo ustvarjalec dokumentov) ali pa neposredno po izročitvi dokumentov in jo izvaja pristojna arhivska inštitucija. V obeh primerih gre za pristop kasnejše pretvorbe, ki smo jo imenovali *zakasnela pretvorba*. Ko so e-dokumenti v arhivski inštituciji, je ta nad njimi prevzela skrbništvo. To pa pomeni tudi prevzem odgovornosti za zagotavljanje dostopnosti in verodostojnosti teh dokumentov ter za njihovo nadaljnjo uporabo. Zato morajo tudi arhivske inštitucije predvideti metodologijo pretvorbe hranjenih arhivskih e-dokumentov v formate in po postopkih, ki bodo to zagotavljali.





Slika 4: Umestitev postopka pretvorbe v življenjski cikel e-dokumentov s statusom arhivskega gradiva (Vir: Hajtnik, 2016)

Ker dandanes sistemi za zagotavljanje dolgoročne e-hrambe (ne glede nato, ali se ta izvaja pri ustvarjalcu ali v arhivski inštituciji) sledijo predvsem standardu ISO 14721 (referenčnemu modelu OAIS), smo postopke pretvorbe umestili v osnovno shemo tega modela (Slika 5):



Slika 5: Umestitev postopka pretvorbe v model OAIS (Vir: Hajtnik, 2016)

- Faza A (pri ustvarjalcu):
  - takoj ob kreiranju/prejemu e-dokumentov (normalizacija, običajno jo izvajajo lastniki e-dokumentov pri ustvarjalcu),
  - v primeru zastaranja formata, v katerem so trenutno e-dokumenti (kasnejša pretvorba, običajno jo izvajajo skrbniki sistema e-hrambe pri ustvarjalcu).
- Faza B (pri ustvarjalcu):
  - *pred zajemom v AIS*, ki v tem primeru predstavlja sistem dolgoročne e-hrambe ustvarjalca, ali *izročitvijo v pristojno arhivsko inštitucijo* (kasnejša pretvorba, običajno jo izvajajo skrbniki sistema e-hrambe pri ustvarjalcu).
- Faza C (pri ustvarjalcu ali v pristojni arhivski inštituciji):
  - pretvorbe se izvajajo neposredno *po prevzemu v AIS*, ki je namenjen dolgoročni e-hrambi; do teh pretvorb pride v primeru, da v prejšnjih fazah pretvorbe niso bile izvedene ali pa je format postal neprimeren za dolgoročno hrambo (primer zastaranja); pretvorbo običajno izvajajo skrbniki AIS (pri ustvarjalcu ali pa v pristojni arhivski inštituciji).

- Faza D (pri ustvarjalcu ali v pristojni arhivski inštituciji):
  - v AIS v primeru zastaranja formatov zaradi potrebe po nadaljnji uporabi e-dokumentov; pretvorbo običajno izvajajo skrbniki AIS (pri ustvarjalcu ali pa v pristojni arhivski inštituciji).

## 6. STANDARD ISO 13008 - OSNOVNI OKVIR ZA IZVAJANJE POSTOPKA PRETVORBE

Postopek pretvorbe teoretično predstavi leta 2012 objavljeni standard ISO 13008, ki pa je bil, čeprav v praksi še ni prav razširjen, v Sloveniji leta 2013 sprejet kot nacionalni standard (SIST ISO 13008).

Ključne aktivnosti, ki jih izpostavlja standard ISO 13008 za izvajanje procesa pretvorbe, so *načrtovanje* (angl. *planning*), *testiranje* (angl. *testing*), *pretvorba* (angl. *conversion*), *validacija* (angl. *validation*), *odobritev* (angl. *sign-off*) in *dokumentiranje* (angl. *documentation*) (slika 6).



Slika 6: Standard ISO 13008 (Vir: Hajtnik, 2016)

Aktivnosti, kot jih predlaga standard, so v nadaljevanju specializirane in podprte s konkretnimi rešitvami (Slika 8).



V fazi **NAČRTOVANJA** se pripravi celoten načrt za uspešno izvedbo pretvorbe, ki naj vključuje naslednje aktivnosti:

### DOLOČITEV OBSEGA

Določiti je treba, kateri e-dokumenti bodo predmet pretvorbe, ter v ta namen pripraviti njihov natančen seznam (evidenco).

## ANALIZA IZBRANEGA OBSEGA E-DOKUMENTOV

E-dokumente, ki bodo predmet pretvorbe, je treba ovrednotiti, analizirati in urediti. V ta namen se izvede:

### Izločitev dokumentov

Iz nadaljnjega postopka se *izloči tiste dokumente*, ki jih ni treba pretvarjati. Razlogi za takšno odločitev so lahko različni, npr.: potekel je rok hrambe, so delovne različice in ne predstavljajo izvirnikov, ki jih je treba ohranjati.

### Preverjanje formatov izvirnih e-dokumentov

*Preveriti je treba, v katerih formatih so dejansko izvirni e-dokumenti*. To se lahko izvede s pomočjo orodja za identifikacijo formatov. Po identifikaciji formatov se e-dokumente istega tipa (npr. besedilni dokumenti, preglednice) razvrsti v skupine dokumentov istega formata.

### Analiza bistvenih lastnosti izvirnikov

Za vsako skupino e-dokumentov posebej se izvede analiza njihovih bistvenih lastnosti, in sicer glede na vsebino, strukturo, kontekst, videz in obnašanje, ter poišče morebitne posebnosti, npr. vključevanje povezav z drugimi dokumenti, posebnosti oblikovanja, vključevanje tabel, grafov, varnostnih vsebin, formul, makrojev, vgrajenih dokumentov, spletnih povezav.

## DOLOČITEV ZAHTEV ZA OHRANJANJE BISTVENIH LASTNOSTI E-DOKUMENTOV IN TEHNIČNIH METAPODATKOV

Določiti je treba *zahteve* za ohranjanje bistvenih lastnosti in tehničnih metapodatkov za vsako skupino e-dokumentov, ki so predmet pretvorbe. Sprejeti je treba odločitev, katere so tiste bistvene lastnosti in tehnični metapodatki, ki jih je treba ohraniti za vsako skupino e-dokumentov, ki so predmet pretvorbe. Tu se lahko uporabi model **MNačel** (Slika 2).

*[Primer: če imajo e-dokumenti arhivsko vrednost, je treba odločitev o določitvi bistvenih lastnosti e-dokumentov, ki ji je treba ohraniti in s tem povezanimi metapodatki, sprejeti skupaj s pristojnim arhivom.]*

## SPREJEM ODLOČITVE, KDO BO IZVEDEL PRETVORBO

Sprejeti je treba odločitev, kdo bo izvedel pretvorbo, sami ali zunanji izvajalec. Da bo proces pretvorbe potekal po načrtih, je treba:

- *identificirati in določiti odgovorne osebe* za izvedbo celotnega procesa pretvorbe, ki imajo dovolj znanja in kompetenc;
- *usposobiti osebe*, vključene v proces pretvorbe: v kolikor se ugotovi, da so potrebna dodatna znanja, je treba zagotoviti dodatno usposabljanje.

[Primer: če so predmet pretvorbe arhivski e-dokumenti, je treba v primeru najema zunanje izvajalca upoštevati tudi predpise. Ti določajo, da mora biti storitev pretvorbe certificirana pri Arhivu RS (ZVDAGA, 72. člen.)]

### **IZBIRA PRISTOPA K PRETVORBI**

Izbira ustreznega pristopa k pretvorbi mora izhajati iz analize e-dokumentov. Glede na ugotovitve iz te faze se določi pristop k pretvorbi glede na:

- obseg in zahtevnost – odločitev izvira npr. iz pomembnosti in vrednosti dokumentov in formata, v katerem so izvirni dokumenti;
- način izvedbe.

### **IZBIRA FORMATA**

Za vsako skupino izvirnih e-dokumentov je treba določiti ciljni format. Ta mora imeti lastnosti formatov, ki so primerni za dolgoročno hrambo.

Format se lahko izbira glede na kriterije, predlagane z modelom **MkiFormat** (Slika 3).

### **IZBIRA PRETVORNIKA**

Pretvornik se lahko izbira glede na predlagane kriterije v modelu **MkiPretvornik** (Slika 7) (Hajtnik, 2016), ki je bil izdelan na podlagi rezultatov raziskav, predvsem pa njihove praktične uporabe pri študiji primerov pretvorb. Analiza skladnosti pretvornikov s predlaganim modelom je podprta s primerom **kontrolnega seznama** (Hajtnik, 2016, priloga 2).

Sprejeti je treba tudi odločitev, kakšni so zahtevani oz. dovoljeni parametri za nastavitve pretvornika.



Slika 7: Model MkiPretvornik - kriteriji za izbiro pretvornika (Vir: Hajtnik, 2016)

## NAČRT IZVEDBE PRETVORBE

Končni izdelek faze načrtovanja je načrt izvedbe, ki dokumentirano opredeljuje, kako morajo potekati posamezne aktivnosti v procesu pretvorbe, katere kontrole je treba izvesti in kako bo celoten postopek pretvorbe dokumentiran.



V fazi **TESTIRANJA** se pripravi celovit načrt, saj je od njegove kvalitete in izvedbe odvisna tudi uspešnost kasnejše pretvorbe. S testiranjem je treba za vsako skupino e-dokumentov ugotoviti, do kakšnih težav ali morda izgub njihovih lastnosti lahko pride pri pretvorbi. Na podlagi tega je treba določiti, katere izgube so sprejemljive in katere ne.

[Primer: pretvorba dokumenta formata DOCX v format .TIFF lahko spremeni številčenje strani. Treba je sprejeti odločitev, ali je ta sprememba sprejemljiva ali ne].

## NAČRT TESTIRANJA

Načrt testiranja mora izhajati iz namena, zakaj izvajamo pretvorbo (npr. zaradi zastaranja formata in prenosa e-dokumentov v AIS ali morda izročitve v pristojno arhivsko inštitucijo).

- **PRIPRAVA NABORA (VZORCA) E-DOKUMENTOV**

Za testne namene je treba določiti velikost testnega vzorca. Če imamo več skupin e-dokumentov (npr. različnih formatov ali glede na tip), je treba za vsako skupino določiti velikost testnega vzorca. Njegova velikost se določi na podlagi predhodno izdelane ocene tveganja, ki predstavlja oceno škode in posledic, če bi pri pretvorbi prišlo do izgube bistvenih lastnosti izvornikov ali njihovega dela. Glede na oceno tveganja lahko testni vzorec predstavljajo vsi ali le izbrani nabor e-dokumentov iz posamezne skupine. V primeru, da bo testni vzorec predstavljal le nabor e-dokumentov, je pomembno, da se pri oblikovanju vzorca določi tudi delež od vseh dokumentov, ki bo preverjen, in ne npr. le število dokumentov. Vzorec mora temeljiti na vključitvi vseh bistvenih lastnosti e-dokumentov (vsebina, struktura, izgled in obnašanje, kontekst) ter na metapodatkih, ki so z njimi povezani. Velikost vzorca lahko niha glede na pomembnost e-dokumentov, kar pa naj bo opredeljeno z oceno tveganja (Hajtnik et al., 2009).

Če je le mogoče, je treba z »lastniki« dokumentov oz. njihovimi primarnimi uporabniki vnaprej določiti tudi merila oz. kontrole, ki bodo omogočile preverjanje in potrditev uspešnosti pretvorbe.

Pomemben del opisa vsake od kontrol je način ukrepanja ob ugotovitvi nepravilnosti. Ta mora vsebovati ne le zahtevo po ponovitvi postopka, pač pa glede na okoliščine tudi preverjanje morebitnih sistemskih napak, kot so: napačna izbira formata, napačna izbira pretvornika ali njegovih nastavitev, nepopolna ali neustrezna navodila, okvara opreme in podobno. Ob odkritju sistemskih napak je treba preveriti napake na vseh e-dokumentih posamezne skupine, pri katerih bi se lahko pojavile, a to ni bilo pred tem zaznano.

- **PRIPRAVA USTREZNEGA INFORMACIJSKEGA OKOLJA (TESTNEGA)**

Vključuje pripravo celotnega informacijskega okolja, v katerem bo mogoče izvesti pretvorbo: strojno in programsko opremo, namestitve in ustrezne nastavitve pretvornikov glede na predhodno določene parametre.

- **TESTNE PRETVORBE IN ANALIZA REZULTATOV**

Izvesti je treba kontrole za preverjanje uspešnosti testne pretvorbe skladno s predhodno določenim načrtom. Kontrole je mogoče izvajati z uporabo metode opazovanja (vizualna kontrola), vendar je v praksi to dokaj nerealen pristop, zato je primernejša uporaba orodij za avtomatizirano preverjanje pretvorbe, če le obstajajo. Tudi ta orodja morajo biti preverjena [*primer orodja za preverjanje pravilnosti pretvorbe v format PDF je veraPDF (Open Preservation Foundation, 2015)*].

Analiza pretvorjenih e-dokumentov mora potrditi uspešnost (popolnost, točnost) pretvorbe v skladu z vsemi predhodno pripravljenimi načrti. Če ugotovimo nepričakovana odstopanja od načrtov, je treba to dokumentirati in analizirati rezultate.

Na podlagi tega se nato preveri, ali so odstopanja sprejemljiva. V kolikor niso, se postopek pretvorbe ne izvede in se poiščejo druge rešitve (npr. drug format, drug pretvornik, drugačen pristop glede na obseg in zahtevnost pretvorbe).

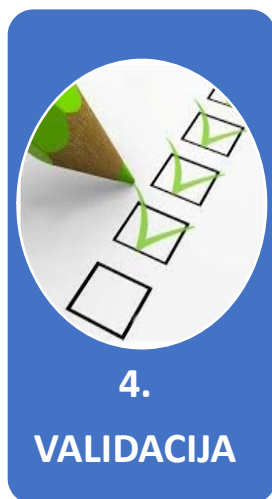
- **DOKUMENTIRANJE POSTOPKA TESTIRANJA:**

Vse izvedene aktivnosti v fazi testiranja je treba dokumentirati.



### FAZA 3: PRETVORBA

Po uspešno izvedeni fazi testiranja se izvede pretvorba izvirnih e-dokumentov v izbrani format skladno z načrti in ugotovitvami v fazi testiranja.

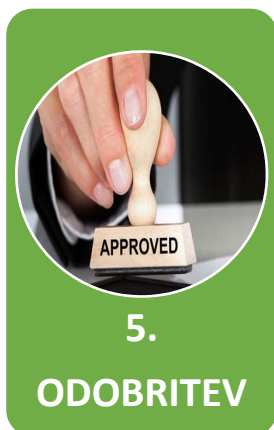


### FAZA 4: VALIDACIJA

Po pretvorbi je treba izvesti analizo rezultatov in kontrolo uspešnosti oz. točnosti pretvorbe. Kontrole se izvedejo skladno s predhodno sprejetimi načrti, vzorec se lahko uporabi enak kot v fazi testiranja. V kontrolo je treba vključiti tudi poročila, ki so jih ustvarili pretvorniki v fazi pretvorbe, oz. vso ostalo dokumentacijo, ki je nastala v procesu pretvorbe.

V fazo validacije je treba vključiti pooblaščen osebe, ki so na koncu odgovorne, da potrdijo uspešnost in točnost pretvorbe. To je pomembno zaradi kasnejšega zaupanja s strani tretjih strank v verodostojnost pretvorjenih e-dokumentov.





## FAZA 5: ODOBRITEV

Po pretvorbi in validaciji morajo pooblaščenec osebe dokumentirano potrditi uspešnost in točnost pretvorbe.

*[Opomba: če so bili predmet pretvorbe arhivski e-dokumenti in se pretvorba izvaja pri ustvarjalcu, naj v fazah validacije in odobritve sodeluje tudi pristojna arhivska inštitucija.]*

## FAZA 6: HRAMBA V AIS

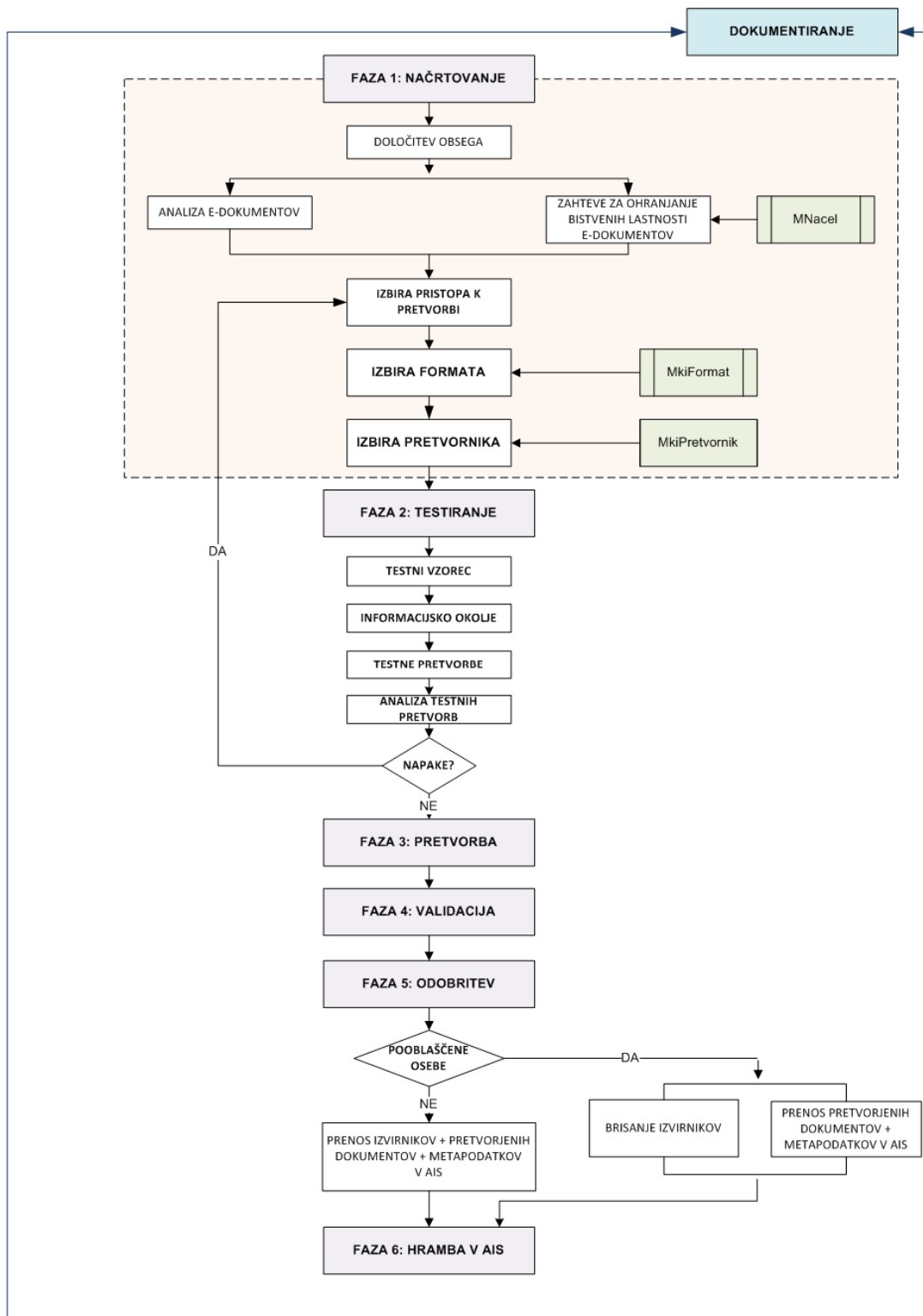
Po uspešno zaključenem postopku pretvorbe je treba pretvorjene e-dokumente shraniti v AIS ter jim tudi določiti rok nadaljnje hrambe. V AIS se shranijo pretvorjeni e-dokumenti (ti so zdaj v izbranem formatu za dolgoročno hrambo) in metapodatki v ločenem dokumentu XML. Po pretvorbi je treba sprejeti tudi odločitev, ali se izvirniki ohranijo ali ne, na kar pa vpliva več dejavnikov. Anketa, izvedena med nacionalnimi arhivi držav članic EU (Hajtnik, 2016, poglavje 3.3), je pokazala, da 91,67 % anketirancev izvirnike po pretvorbi ohranja. Na to odločitev pomembno vplivajo tudi predpisi (Hajtnik, 2016, poglavje 3.5.2). Ključno pri odločitvi pa je predvsem to, kako uspešna in točna je bila pretvorba.

## 7. DOKUMENTACIJA

Dokumentiranost celotnega procesa pretvorbe, kar vključuje vse faze in aktivnosti, je ključna za zagotavljanje nadaljnje varne hrambe pretvorjenih dokumentov. Pri tem ima posebno vlogo dokumentiranje vseh kontrol in aktivnosti, izvedenih po njih.

Dokumentiranje postopkov je lahko izvedeno ročno ali strojno. Pri ročnih kontrolah so najbolj običajne oblike dokumentiranja zapisniki, poročila in različne evidence. Pri strojno podprtem dokumentiranju nastajajo različne oblike zapisov, od dnevniških zapisov (logov) do revizijskih sledi. Te lahko v določenih primerih v celoti nadomestijo ročne kontrole.

Dokumentacija o izvajanju postopka pretvorbe naj se hrani enako časovno obdobje, kot velja za hrambo e-dokumentov. Ta čas bi se lahko spremenil v primeru, da je bila izvedena posebna revizija postopka pretvorbe s strani pooblaščenih oseb (npr. zunanji revizijski pregled), ki je potrdila točnost pretvorbe, s tem pa tudi verodostojnost pretvorjenih e-dokumentov.



Slika 8: Celovit pristop k pretvorbi

## 8. ZAKLJUČEK

Standard ISO 13008 predstavlja dober teoretičen okvir za izvajanje postopka pretvorbe e-dokumentov v druge formate. Specializiran s konkretnimi rešitvami, predvsem z modelom kriterijev za izbiro ustreznih formatov *MkiFormat* in modelom kriterijev za izbiro pretvornikov *MkiPretvornik*, pa predstavlja celovit pristop k pretvorbi, s katerim lahko zmanjšamo izgubo bistvenih lastnosti e-dokumentov ob njihovi pretvorbi ter s tem povečamo možnosti ohranjanja njihove verodostojnosti in ponovne uporabe za ves čas hrambe.

## LITERATURA

- ALCTS - Association for Library Collections & Technical Services. (2007). *Definitions of digital preservation*. Chicago: ALA Annual Conference, Washington. Najdeno 17. 1. 2016 na naslovu <http://www.ala.org/ala/mgrps/divs/alcts/resources/preserv/defdigpres0408.cfm>.
- Bearman, D. (1999). Reality and Chimeras in the Preservation of Electronic Records. *D-Lib Magazine* 5 (4), str. 6–11. DOI: 10.1045/april99-bearman.
- Cox, R. J. (2001). *Managing records as evidence and information*. West-port, Connecticut: Quorum Books, London, str. 243.
- Duranti, L. (1998). *Diplomatics: New Uses for an Old Science*. Scarecrow Press, str. 186.
- Duranti, L. (2001a). The impact of digital technology on archival science. *Archival Science, Vol 1*, str. 39–55. DOI: 10.1007/BF02435638.
- Duranti, L. (2001b). Concepts, Principles, and Methods for the Management of Electronic Records. *The Information Society*, vol. 17 (4), str. 271–279. DOI:10.1080/019722401753330869.
- Enotne tehnološke zahteve 2.1. I.del: *Uvodna pojasnila /ETZ 2.1, I. del/*. (2013). Arhiv RS. Najdeno 15. 11. 2015 na naslovu [http://www.arhiv.gov.si/si/zakonodaja\\_in\\_dokumenti/](http://www.arhiv.gov.si/si/zakonodaja_in_dokumenti/).
- File-Extensions.org. (2015). *The Source for File Extensions Information*. Najdeno 25. 11. 2015 na naslovu <http://www.freefileconvert.com/>.
- Hajtnik, T. (2012). Analiza primernosti formata PDF/A za dolgoročno hrambo gradiva v digitalni obliki. *Arhivi*, letn. 35 (1), str. 157–182.
- Hajtnik, T. (2016). *Celovit pristop k pretvorbi elektronskih dokumentov v obliko za dolgoročno hrambo : doktorska disertacija*. Maribor: [T. Hajtnik], 243 str.
- Hajtnik, T. (ur.), Škofljanec, J., Mrdavšič, A., Zupančič, M., Vodopivec, J., Dobernik, M. (2009). *Arhiviranje, hramba in upravljanje dokumentov: Navodila in praktični primeri pravilne hrambe vseh vrst dokumentarnega gradiva za podjetja in ustanove*. Maribor: Založba Forum Media, mapa, nevezani listi, periodične, trimesečne posodobitve.
- Hedstrom, M. & Montgomery, S. (1998). *Digital Preservation Needs and Requirements in RLG Member Institutions*. A study commissioned by the Research Libraries Group. California, USA, str. 37.
- International Council on Archives (ICA) - Committee on electronic records. (1997). *Guide for managing electronic records from an archival perspective*. Paris, str. 58.
- IDC. (2014). *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*. Pridobljeno s <https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm>.
- ISO 13008: 2012. *Information and documentation - Digital records conversion and migration process*. International Organization for Standardization. Leta 2013 sprejet kot slovenski

nacionalni standard SIST ISO 13008:2013 Informatika in dokumentacija - Proces konverzije in migracije digitalnih zapisov.

- ISO 14721:2012. *Space data and information transfer systems -- Open archival information system - Reference model*. International Organization for Standardization.
- ISO 15489-1: 2001. *Information and documentation - Records management - Part 1: General*.
- ISO 23081-1: 2006. *Information and documentation - Records management processes - Metadata for records - Part 1: Principles*.
- ISO 30300: 2011. *Information and documentation — Management systems for records — Fundamentals and vocabulary*.
- Open Preservation Foundation. (2015). *veraPDF 0.8 now available*. Najdeno 25.12.2015 na naslovu <http://openpreservation.org/news/verapdf-0-8-now-available/>
- Puglia, S. (1999). Creating Permanent and Durable Information: Physical Media and Storage Standards. *CRM (Cultural Resource Management)*, 22 (2), str. 25–27.
- Rothenberg, J. (1999). Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation (*A Report to the Council on Library and Information Resources*). Washington: Council on Library and Information Resources, DC, str. 41.
- Rothenberg, J., & Bikson, T. K. (1999). Carrying authentic, understandable and usable digital records through time. *Technical report, Report to the Dutch National Archives and Ministry of the Interior*. RAND-Europe, str. 114.
- SSKJ - *Slovar slovenskega knjižnega jezika*, spletna izdaja. (2000). SAZU in ZRC SAZU, Inštitut za slovenski jezik Frana Ramovša in avtorji. Založba ZRC. Najdeno 1.12.2015 na naslovu <http://bos.zrc-sazu.si/sskj.html>
- The National Archives. (2011). *The technical registry PRONOM*. Najdeno 7.11.2011 na naslovu <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>.
- Thomas, S. & Martin, J. (2006). Using the papers of contemporary British politicians as a testbed for the preservation of digital personal archives. *Journal of the Society of Archivists*, 27(1), str. 29-56. DOI: 10.1080/00039810600691254
- Thomas, S. & Martin, J. (2006). Using the papers of contemporary British politicians as a testbed for the preservation of digital personal archives. *Journal of the Society of Archivists*, 27(1), str. 29-56. DOI: 10.1080/00039810600691254
- Thomassen, T. (2001). A First Introduction to Archival Science. *Archival Science*, vol. 1, str. 373–385.
- Waters, D., & J. Garrett (1996). *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information commissioned by the Commission on Preservation and Access and the Research Libraries Group*, Washington, D.C.: Commission on Preservation and Access, str. 71.
- Zakon o varstvu dokumentarnega in arhivskega gradiva ter arhivih /ZVDAGA/. *Uradni list RS*, št. 30/06, 24/14 - Odl. US in 51/14.

## SUMMARY

### CONVERTING ELECTRONIC DOCUMENTS INTO FORMATS FOR LONG-TERM PRESERVATION ON THE BASIS OF STANDARD ISO 13008

Conversion of digital documents into formats suitable for long-term preservation of such documents is currently the most widespread strategy for long-term preservation of digital records. However, when taking into consideration the fact that access to and use of digital documents depends mostly on the ever-changing technology and that currently adopted standards are not yet stable enough and without deficiencies, one has to approach the process of conversion holistically and with caution. Theoretical framework for doing so is ISO 13008 standard, which in places is fully specified by the author and supported by concrete solutions. Here we should point out the importance of criteria needed to define a certain preservation of digital records as safe. Such criteria are presented in the form of **MNačel** model which includes five basic principles: accessibility, usability, integrity, authenticity and re-use of digital documents throughout the whole period of their preservation. For digital documents to retain their main properties during the process of conversion, it is essential that suitable formats and converters are chosen. To select long-term preservation suitable formats the author suggests the application of criteria model **MkiFormat**, whereas for the selection of suitable converters the model **MPretvornik** is suggested. By doing so, we are building a holistic approach to digital documents conversion which can eventually increase the chances of safe long term preservation of digital documents and their re-use.