

# Iz prakse za prakso

1.01 Izvirni znanstveni članek

UDK 657.471:004:930.25

Prejeto: 12. 1. 2006

## Stroškovni model digitalne hrambe\*

**JACQUELINE SLATS**

vodja oddelka za dolgoročno digitalno hrambo

Nacionalni arhiv Nizozemske, Prins Willem Alexanderhof 20, 2595 BE Den Haag, Postbus 90520,  
2509 LM Den Haag

el. pošta: jacqueline.slats@nationalearchief.nl

**REMCO VERDEGEM**

višji svetovalec za dolgoročno digitalno hrambo

Nacionalni arhiv Nizozemske, Prins Willem Alexanderhof 20, 2595 BE Den Haag, Postbus 90520,  
2509 LM Den Haag

el. pošta: remco.verdegem@nationalearchief.nl

### IZVLEČEK

Projekt 'Digital Preservation Testbed' je preučil stroške za dolgoročno digitalno hrambo, pripravil seznam stroškovnih indikatorjev, ki vplivajo na celotne stroške hrambe, in razvil model v excel-u za izračun celotnih stroškov za hrambo pa tudi za primerjavo stroškov za različne načine hrambe, ki jih je raziskal 'Testbed'. Stroškovni model je namenjen nacionalnim ali regionalnim arhivskim ustanovam, možno pa ga je uporabiti tudi za potrebe lokalne hrambe v ministrstvih ali drugih državnih organih. Stroški, opisani v članku, bodo vedno upoštevani, ne glede na to, ali bodo dokumenti hranjeni trajno ali pa samo 10 do 20 let. Raven hrambe ali digitalni arhivski sistem ter relativna velikost komponent instalacije se lahko spreminjajo, vendar je za vsak primer potrebno posebej preučiti vse opisane dejavnike. Prav tako bi morali biti že na začetku v model vključeni tudi stroški za prihodnje spremembe. Hramba se bo v naslednjih generacijah razvijala in spreminjala, prav tako pa tudi oprema za hrambo.

KLJUČNE BESEDE: digitalna hramba, dolgoročna hramba, stroškovni model, indikatorji stroškov

### ABSTRACT

#### COST MODEL OF LONG TERM DIGITAL PRESERVATION

The Digital Preservation Testbed has studied the costs of long term digital preservation, has prepared a list of the cost indicators that influence the total cost of preservation, and has developed a model in Excel for calculating the total cost of preservation, and for comparing the costs of the different preservation approaches Testbed has investigated. The cost model is designed around a national or regional archival institution, but it is also applicable to the local storage needs of a ministry or of any other governmental agency. Whether records are being kept forever, or just for 10 to 20 years, the costs described in the paper will be incurred. The scale of the storage or digital archive system, and the relative sizes of the different parts of the installation may vary, but all of these factors need to be considered in each case. Costs for the future changes should be built into the model from the start. Preservation is going to evolve and change over the next generation, and a preservation facility will have to change too.

KEY WORDS: digital preservation, long term preservation, the cost model, cost indicators

\* Prispevek je bil predstavljen na konferenci DLM Forum 5.–7. oktobra 2005 v Budimpešti.

## 1. Uvod

Projekt z naslovom "Digital Preservation Testbed" je bil praktični raziskovalni projekt, katerega glavni cilj je bil raziskati možnosti za dolgoročno varovanje sprejemljive dostopnosti do avtentičnega arhivskega gradiva z izvajanjem eksperimentov v nadzorovanem in varnem okolju.

Tako so bili v okviru projekta dognani učinki izvajanih aktivnosti za hrambo pri različnih arhivskih dokumentih.

Projekt Testbed je raziskal tri različne načine dolgoročne digitalne hrambe: migracijo, XML ter emulacijo (bolj natančno: način UVC [Universal Virtual Computer]). Ovrednotena je bila ne samo učinkovitost vsakega načina, ampak tudi njihove omejitve, aplikativne možnosti ter stroški.

Poizkusi so potekali na štirih različnih vrstah dokumentov: besedilni dokumenti, preglednice, sporočila elektronske pošte ter podatkovne zbirke različnih velikosti, kompleksnosti in značilnosti.

Projekt "Digital Preservation Testbed" je spomladi 2005 strnil svoje ugotovitve:

- svetovanje, kako ukrepati s tekočimi digitalnimi dokumenti,
- priporočila za primeren način hrambe ali oz. kombinacija različnih načinov za posamezne vrste dokumentov,
- funkcionalne zahteve za sistem hrambe,
- model odločitev za izbiro pravilne strategije hrambe,
- priporočila za arhivska navodila in predpise,
- stroškovni indikatorji in stroškovni model za digitalno hrambo.

Referat opisuje različne indikatorje stroškov digitalne hrambe in se osredotoča na primerjavo stroškov za različne načine hrambe, ki so jih preučili v okviru projekta "Testbed". Oblikovan je bil seznam kazalcev, ki kažejo vpliv na celotni strošek hrambe. Razvit je bil tudi ocenitveni model v excelu za izračun celotnih stroškov hrambe ter za primerjavo stroškov pri uporabi različnih načinov hrambe.

Ti stroški so ocene, ki temeljijo na študijah in izkušnjah v okviru projekta "Testbed", objavljenih informacijah ter informacijah, ki so jih drugi prispevali za projekt, kot tudi na osnovi 'splošne presoje'. S podrobnimi ocenami smo želeli spodbuditi tudi druge, da bi posredovali svoje mnenje o teh številkah in poročali o resničnih stroških, ki nastanejo v praksi.

Kazalci stroškov, ki so določeni, bodo vedno upoštevanji, ne glede na to, ali morajo biti določeni dokumenti hranjeni največ 10–20 let ali pa pridejo

v poštev za trajno hrambo. Četudi se lahko nekateri elementi spreminjajo, in sicer raven digitalnega arhivskega sistema, sistem digitalne hrambe ter relativna velikost komponent instalacije, pa bo treba stroškovne dejavnike (opisane v nadaljevanju) v vsakem primeru upoštevati.

Četudi na prvi pogled kaže, da je seznam zelo podroben, pa je vendarle pomembno, da ne spregledamo nobenega od stroškovnih kazalcev. Še posebej bo treba izračunati investicijske stroške in stroške za osebje. Digitalna hramba se bo kontinuirano razvijala in spreminjala. Zato se bo morala spreminjati tudi funkcionalnost za sprejemljivo hrambo digitalnih dokumentov. Stroški, povezani z izvajanjem sprememb v bodoče, morajo biti že v samem začetku vključeni v cenitveni model.

## 2. Kazalci stroškov

"Testbed" razlikuje med sledečimi kazalci stroškov:

- a) stroški za digitalni arhivski sistem (= digitalno skladišče) in funkcionalnost za dolgotrajno hrambo digitalnih dokumentov (= sistem hrambe),
- b) stroški za osebje,
- c) stroški za razvoj ali nakup programske opreme in metod/strategij za hrambo digitalnih dokumentov,
- d) stroški za dejansko izvajanje določenih aktivnosti za hrambo,
- e) drugi faktorji, ki vplivajo na celotne stroške hrambe.

### a) *Stroški za digitalni arhivski sistem in funkcionalnost za dolgotrajno hrambo digitalnih dokumentov*

Stroški za digitalno skladišče in sistem za hrambo so sestavljeni iz različnih komponent. Stroškovni model prikazuje tako glavne kot tudi manjše dejavnike. Model tudi pojasnjuje, na katere kazalce vplivajo različni načini nastanka dokumentov in različne strategije hrambe ter kateri kazalci niso občutljivi za (strateško) izbiro te narave.

#### • *Fizični prostor*

Fizični prostor je zahtevan za sisteme za hrambo in dolgotrajno hrambo. Strežniki bodo potrebni za skladiščenje digitalnih dokumentov in upravljanje dolgoročne hrambe. Svetujemo lahko postavitev ločene opreme za razvoj, testiranje in produkcijo dolgoročne hrambe.

- *Strojna oprema za digitalni arhivski sistem (npr. strežniki, mediji za shranjevanje, oprema za kopiranje)*

Za shranjevanje dokumentov je potrebna strojna oprema (v datotečnem sistemu, v arhivskem skladišču ali v RMA).<sup>1</sup> Ko bo pridobljena ocena o številu dokumentov, ki jih bo potrebno hraniti, bo potrebno tudi konfigurirati opremo za skladiščenje, četudi bo težko predvideti obseg prenosov v prihodnjih letih.

- *Programska oprema za digitalni arhivski sistem (npr. OS,<sup>2</sup> programske aplikacije, varnostno programje, programje za prikaz, programje komunikacij)*

To zajema vprašanja, kot so nakup operacijskih sistemov in podatkovnih zbirk standardne programske opreme. Prav tako bo potrebno programje za varnost (zoper viruse, nepooblaščen dostop ter nedovoljeno spreminjanje dokumentov s strani nepooblaščenih oseb). Mogoče bo treba imeti tudi posebno programsko opremo za pridobivanje in hrambo arhivskih dokumentov v digitalni obliki.

- *Strojna oprema za sistem hrambe (npr. strežniki, delovne postaje, mediji za shranjevanje, oprema za kopiranje)*

Sistem za hrambo lahko zahteva isto vrsto in velikost računalniškega sistema (strežniki in skladišča) kot digitalni arhivski sistem. To je potrebno zato, da lahko sistem hrambe sprejme skupine dokumentov s skupno velikostjo, ki presega nekaj terabajtov. Sistem bo moral hraniti te dokumente na varen način ter pripravljene za izvajanje operacij za hrambo (kot so migracija ali konverzija v XML), ocenitev rezultatov operacij hrambe in vrnitev dokumentov v digitalni arhivski sistem.

- *Programska oprema za sistem hrambe (npr. OS, programske aplikacije, varnostno programje, orodja za hrambo, programje za testiranje in ocenjevanje, komunikacijsko programje)*

Sistem za hrambo lahko zahteva več kot en operacijski sistem, ker je lahko to potrebno za prenos dokumentov iz njihovega prvotnega operacijskega sistema v drug operacijski sistem, ki je sposoben omogočiti izboljšano hrambo. Poleg tega se lahko pokaže tudi potreba po več kot enem programskem okolju, če organizacija načrtuje razvoj lastnega (hišnega) programskega orodja ali prilagodi orodje nekoga tretjega.

## **b) Stroški za osebje**

Ta sekcija daje pregled obveznosti zaposlenih, ki so vključeni v digitalni arhivski sistem ter v sistem za hrambo. V diskusiji sta razvidna število in vrsta zaposlenega osebja, ki bo potrebno. Stroškovni model, razložen pozneje v tem poglavju, temelji na času, ki bo potreben temu osebju, le-to pa ima različne kvalifikacije in sposobnosti.

Stroški za osebje so vedno glavni dejavnik. Potrebno osebje lahko dobimo s selekcijo ali posebnim novačenjem za delo na digitalnem arhivu in sistemih za hrambo. Vendar bi bilo v nekaterih primerih zaželeno, da bi pridobili tudi pomožno – dodatno osebje iz drugih disciplin (npr. iz oddelka za upravljanje dokumentov ali ICT oddelka).

- *Osebje digitalnega arhivskega sistema*

Osebje bo moralo začeti oblikovati in graditi digitalni arhiv. To bo zahtevalo proračun v obsegu: enega do dveh človek-let. Stroškovni izračuni tega ne bi smeli podcenjevati.

- *Osebje sistema za hrambo*

Osebje, odgovorno za sistem hrambe, bo predvsem moralo tudi oblikovati in zgraditi sistem. Vzpostaviti bo moralo sistem nadzora kakovosti, SOP-se<sup>3</sup> ter postopke. Končno bo moralo začeti razvijati metode hrambe in ocenitvene teste ter začeti izvajati sprejemljivo hrambo dokumentov. Naj poudarimo še enkrat, da stroške, vključene v fazo razvoja in izgradnje zelo hitro podcenijo.

- *Osebje javne službe*

Osebje, ki je odgovorno za pomoč strankam pri dostopanju do "odprtih" dokumentov.

## **c) stroški za razvoj (ali nakup) programske opreme in metod za hrambo dokumentov**

Eden od zaključkov projekta "Testbed" je bil, da digitalna hramba ni vprašanje 'vse ali nič'. V mnogih primerih je mogoče lastnosti dokumentov, ki so bistvene za celovitost in verodostojnost, ločiti od drugih, manj pomembnih lastnosti. Zato se lahko aktivnosti digitalne hrambe osredotočijo le na te bistvene vidike celovitosti in verodostojnosti dokumenta.

Prvotni stroški, vključeni v digitalno hrambo, se nanašajo na vprašanja, kot je določanje zahtev za verodostojnost za vsako serijo dokumentov. V idealnih razmerah bodo te zahteve določili upravjalci

<sup>1</sup> RMA (return merchandise authorization), op. prev.

<sup>2</sup> OS – operacijski sistem, op. prev.

<sup>3</sup> Standardni operativni postopek, op. prev.

dokumentov. Vendar pa bo v nekaterih okoliščinah zahteve za verodostojnost morala določiti multidisciplinarna skupina, sestavljena iz strokovnjakov, kot so upravljalci arhivskega gradiva, specialisti za IT; pri tem bi moral vsak član skupine imeti specifična znanja s svojega področja.

Stroškovni model predpostavlja, da bodo zahteve (za verodostojnost) določene za vsako serijo dokumentov. Ena serija vsebuje samo dokumente, v kateri so vsi narejeni z isto aplikacijo, njihov prevzem ali hramba pa se izvede hkrati oziroma istočasno. V nadaljevanju bo prikazano, da je velikost serije pri stroških odločilen dejavnik.

Ko so zahteve za verodostojnost enkrat določene, je naslednji korak oblikovati in razviti primeren način hrambe. Ker je to dolgotrajen proces, ki zahteva več raznih veščin (izkušenj in znanja), predvidevamo, da se bo postopoma razvila celota mednarodnih izkušenj o strategijah hrambe in da si jih bomo lahko delili. Vendar pa bo kljub temu potrebno oceniti te strategije glede na specifične zahteve določene serije dokumentov, ki bo obravnavana. Nazadnje pa bo potrebno v določenih razmerah prilagoditi določen način hrambe.

Končno bo morala biti vsaka strategija ali način testiran in dokumentiran. Vsi postopki IT, ki so vključeni v katerikoli sistem hrambe, se bodo morali podrediti najstrožjim standardom kakovosti. Visoka raven kakovosti je bistvenega pomena za verodostojnost, kajti sistem kakovosti in dokumentacije sta potrebna za dokaz, da so aktivnosti za hrambo dale pričakovane rezultate in niso imele nikakršnega vpliva na druge dokumente. Visoka stopnja kakovosti tudi povečuje možnost, da bo postopek ponovno uporabljen tudi pri drugih sistemih za hrambo.

#### **d) stroški za izvajanje aktivnosti za hrambo**

Ta sekcija daje pregled stroškov, vključenih v izvajanje aktivnosti za hrambo digitalnih dokumentov. Znotraj tega konteksta lahko "izvajanje aktivnosti za hrambo" povežemo z različnimi aktivnostmi:

- migracija dokumentov (transformacija)
- pretvorba dokumentov v XML
- uporaba UVC za ohranitev dostopnosti dokumentov

Te je mogoče pojasniti s terminologijo OAIS.<sup>4</sup> Migracija ali druga oblika pretvorbe dokumentov se kaže v spremembah "Arhivskega informacijskega paketa" (Archival Information Package – AIP), ki

je shranjen v digitalnem arhivskem sistemu. AIP1 se spremeni v AIP2. AIP2 pa je kot osnova za DIP (Dissemination Information Package), ki je izdan proslcem.

Pretvorba v XML je druga oblika pretvorbe, ki jo je "Testbed" preiskoval kot možen način zanesljive hrambe dokumentov. V nadaljevanju so razložene možne stroškovne prednosti XML (ker je odprt standard, pričakujemo, da bo imel daljšo življenjsko dobo, bo uporabnejši in ga bodo lahko interpretirale različne aplikacije). Pretvorba v XML spremeni AIP iz AIP1 v AIPX, AIPX pa se spremeni v XML.

Ohranjanje dostopa do digitalnih dokumentov z uporabo emulacije strojne opreme nima vpliva na dokumente, ki so v AIP. Načeloma bo tudi DIP ostal v prihodnosti nespremenjen, četudi bo v praksi mogoče potrebno narediti več manjših sprememb DIP, da bi ga prilagodili bodočim tehnologijam. V tem pogledu je Testbed raziskal način UVC, ki ga je formuliral IBM. Ta način temelji delno na emulaciji in delno na migraciji.

Kot bo pokazal stroškovni model, predstavljajo stroški za aktivnosti digitalne hrambe (stroški za delovanje) dejansko le manjši delež celotnih stroškov za hrambo. Stroški izvajanja digitalne hrambe so odvisni tudi od velikosti serije: stroškovni model bo razkril, da je glede na stroške ugodneje grupirati dokumente v večje serije.

#### **e) drugi dejavniki, ki vplivajo na celotne stroške hrambe**

Drugi dejavniki, ki sicer niso omenjeni v povzetku, so prav tako zelo pomembni. Vendar pa imajo lahko taki posredni dejavniki znaten vpliv na delež tudi pri celotnih stroških. Poleg tega lahko vplivajo tudi na število omenjenih dejavnikov.

- *Javna storitev*

Velik vpliv na stroške bo imela raven, do katere si lahko uporabniki naložijo storitve arhiva in sistemov za hrambo. Vendar pa "pridržek" take storitve tudi omogoča pridobitev dohodka.

- *Čas med (posameznimi) aktivnostmi hrambe*

To je odločilen stroškovni dejavnik. Več kot je aktivnosti za hrambo, večji so stroški. Ker naraščanje števila aktivnosti hrambe, poleg tega, povečuje tveganje, to pa vpliva na verodostojnost in popolnost (neokrnjenost) dokumentov, bodo zaradi tega potrebna dodatna testiranja.

Stroške je mogoče zniževati v daljšem časovnem obdobju med (posameznimi) aktivnostmi

<sup>4</sup> [http://ssdoo.gsfc.nasa.gov/nost/isoas/ref\\_model.html](http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html)

hrambe. Vendar pa lahko aktivnosti hrambe, ki se izvajajo v prevelikih časovnih intervalih, povečajo tveganje za probleme z digitalno hrambo in stroške hrambe.

- *Spremljanje tehnologije – ocenjevanje, kadar tveganje narasča*

Spremljanje tehnologije zahteva kontroliranje strojne in programske opreme ter sistemov, uporabljenih za veljavne dokumente. Zaradi nevarnega zastarevanja komponent, od katerih so odvisni digitalni dokumenti, bo nastala potreba po tem, da se ovrednotijo in izvedejo potrebni ukrepi.

- *Zahteve za dopolnilno hrambo*

"Testbed" priporoča tudi hrambo izvirnih (originalnih) datotek hranjenih dokumentov. Za hranjenje besedilnih dokumentov priporočamo hrambo v obeh formatih – PDF in XML. Ta priporočila povečujejo prostor, ki je potreben za hrambo vsakega dokumenta. V nekaterih primerih se lahko prostor poveča za tri- do petkrat. Čeprav je hramba razmeroma poceni, pa se to lahko pokaže v dodatnih stroških.

- *Povezave s sistemom za upravljanje elektronskih dokumentov*

"Testbed" ni raziskal povezav v DMS<sup>5</sup> ali RMA.<sup>6</sup> Vendar pa lahko pričakujemo, da bodo te povezave zaželeno v prihodnosti. To bo pomenilo dodatne stroške za oblikovanje in vzdrževanje teh povezav.

- *Obseg dokumentov*

Pričakovani obseg dokumentov za hrambo in upravljanje bo znatno vplival na stroške. Stroški za hrambo naraščajo linearno z obsegom. Poleg tega se bo potreben prostor še hitreje povečeval, če bodo morali biti dokumenti (s)hranjeni v različnih formatih (na primer datoteka v izvirnem formatu ter v dveh migriranih formatih).

Pri obsežnejših dokumentih bodo potrebni dražji strežniki in sistemi za hrambo (več kot 500 terabajtov), še posebej tam, kjer je potreben hiter dostop do dokumentov.

Vendar bi bilo treba upoštevati, da na stroške digitalne hrambe bolj vpliva različnost (raznovrstnost) kot pa obseg dokumentov. Dokumenti, ki uporabljajo različne funkcije neke aplikacije ali različno programsko opremo, bodo na splošno potrebovali različne strategije za hrambo ali vsaj raz-

lična testiranja strategij za hrambo. Iz tega razloga bo stroškovno ceneje hraniti nekaj velikih serij dokumentov, od katerih vsi uporabljajo isto aplikacijo (mogoče tudi enak "template")<sup>7</sup> in imajo vsi enake zahteve za verodostojnost, kot pa hraniti manjše vendar različne serije, ki zasedajo enak obseg prostora za hrambo.

- *Zahteve za verodostojnost in zanesljivost*

Zahteve po verodostojnosti za posebne vrste dokumentov predstavljajo pomemben stroškovni dejavnik. Pretehtajmo na primer besedilni dokument. Njegova hramba bo razmeroma preprosta naloga, če bi bilo treba ohraniti samo besedilo (vsebino) dokumenta. Lahko pa se ohranijo tudi poudarki. Vendar pa bodo stroški naraščali, če bo treba ohraniti natančno lego vsake črke na določeni strani ter natančno barvo. To pa bo otežilo tudi testiranja postopkov hrambe. Zato je tudi pomembno, da so zahteve za verodostojnost določene na kolikor le mogoče najbolj izčrpen in stvaren način.

- *Ohranjanje sistemov samih*

Ne nazadnje bo treba ohraniti tudi sisteme same. Stroške za to bo mogoče deloma pokriti z nižanjem vrednosti (amortizacija), rezultat tega pa bodo razpoložljiva sredstva za obnovo sistemov v tri- do petletnem obdobju. Vendar pa je mogoče, da posebni elementi sistema hrambe tvorijo del digitalnega dokumenta ali objekt hrambe<sup>8</sup>, posledica tega pa je, da bosta morala biti hranjena ločeno. Vsekakor pa izvoz dokumentov iz sistema A v sistem B ne bo enostavno početje.

### 3. Stroškovni model

V tem delu bo pojasnjen stroškovni model. Ta ocenitveni model (verzija 1.0, 20. april 2005), razvit v excelu 97, je dostopen na naši spletni strani

(<http://www.digitaleduurzaamheid.nl/index.cfm?paginakeuze=185&categorie=6>). Sestavljen je iz treh preglednic. Prva ("Cost basis" – stroškovne osnove) je podlaga za drugi dve tabeli, opisuje pa stroške za osebje in digitalni arhivski sistem ter sistem za hrambo. Druga preglednica ("Time calculation" – izračun časa) prikazuje čas, ki je potreben za izvajanje več aktivnosti za vzdrževanje dokumentov, za razvoj načinov hrambe in za dejansko

<sup>5</sup> DMS (Document Management System), op. prev.

<sup>6</sup> Glejte opombo 1.

<sup>7</sup> Template = standardizirani vzorci (vnaprej pripravljene obrazci, predloge), op. prev.

<sup>8</sup> Glejte poglavje 5 (Priporočila za podatkovno zbirko) za nadaljnje razlage pojma "objekt hrambe".

izvajanje aktivnosti za hrambo (po vrstah osebja, na leto in za obdobje N let). Tretja preglednica ("Cost calculation" – izračun stroškov) se posveča primerjavi različnih načinov hrambe, ki jih je projekt Testbed raziskal (migracija [vključno z pretvorbo v PDF]; uporaba XML; ter emulacija [Universal Virtual Computer], in sicer stroškovno (na serijo dokumentov, na leto in za obdobje N let).

Ponovno je treba poudariti, da stroškovni model ni mišljen kot nespremenljiv in končen; njegov namen je predstaviti izhodiščno točko za izboljšave in spremembe, ki temeljijo na skupnih izkušnjah in povratnih informacijah. Prikazani stroški so ocene, ki temeljijo na raznih poskusih in poskusih "Testbed", objavljenih podatkih ter podatkih, ki so jih drugi zagotovili za projekt "Testbed" ter na osnovi 'splošne presoje'. Z objavo teh podrobnih ocen smo želeli spodbuditi druge skupine k razpravi o teh številkah in k temu, da bi poročali o velikosti dejanskih stroškov v praksi.

Naslednja podpoglavja prikazujejo več najpomembnejših zaključkov, ki temeljijo na rezultatih stroškovnega modela. Podpoglavja so urejena takole: Aktivnosti vzdrževanja dokumentov, Razvojne aktivnosti ter Izvajanje digitalne hrambe. Prikazani stroški so izračunani za obdobje od enega do N let. V začetku je opredeljenih nekaj predpostavk, na katerih temelji model ceniitve.

### **Predpostavke, temelječe na cenitvenem modelu**

Predpostavlja se, da obstaja šest kategorij osebja, ki pa je razvrščeno v štiri plačilne razrede (oz. glede na štiri različne urne postavke za plačo). Vsaka kategorija osebja naj bi delala najmanj 1620 ur na leto. Stroškovna kategorija je določena za vsako kategorijo osebja in bo v nadaljevanju uporabljena v preglednicah.

Predpostavlja se, da bosta digitalni arhiv in sistem za hrambo potrebovala štiri vrste prostora. Ocena je bila narejena na podlagi stroškov za opremljanje vseh vrst prostorov. Predlog je zajel tudi število zaposlenih za polni delovni čas, in glede na vrsto prostora v vsaki kategoriji. Naj še enkrat opozorimo: možni so popravki cenitvenega modela, ki bodo vplivali na izračunane rezultate.

Navedena oprema naj bi bila sposobna upravljati z obsegom 100 terabajtov dokumentov ter bi jo bilo mogoče razširiti na upravljanje obsega 1000 terabajtov (1 petabajt) ali več.

Predpostavljeno je, da bo 10 % dokumentov, vstavljenih v digitalni arhivski sistem, potrebovalo poseg (odstranitev gesla, ravnanje z nenavadnimi

fonti, avtomatizirana polja za datum, makroji, itd.).

Predpostavljeno je, da dokumenti, oblikovani neposredno v XML, ne bodo potrebovali posegov ali dodajanja metapodatkov.

Predpostavljeno je, da se bo migracija ponavljala vsaka tri leta, pretvorba v XML (in/ali njegovih naslednikov) pa vsakih deset let. Tukaj ni ponavljajočega se dejavnika za UVC (ter prikazovalnik), ker je predpostavljeno, da bodo prilagoditve za izboljšavo UVC ter strojno opremo prikazovalnika zgolj minimalne.

Za način UVC je predpostavljeno, da bo moral biti za vsako leto razvit program dekodiranja datumskega formata.

### **Aktivnosti za poslovanje z dokumenti**

"Testbed" je raziskal štiri vrste dokumentov (elektronska sporočila, besedilne dokumente, preglednice, podatkovne zbirke) – na podlagi priporočil, ki jih je objavil pa postaja razvidno, da se digitalna hramba začne že v času nastajanja dokumentov. Oblikovanje dokumentov na primeren način je hitrejši, cenejši in manj tvegan način pridobivanja zanesljivih dokumentov v primerjavi s "popravki" teh dokumentov pozneje.

Dober primer teh učinkov je različnost stroškov elektronskih sporočil, ki so bila shranjena iz standardne aplikacije elektronske pošte (kot je outlook), v primerjavi z elektronskimi sporočili, shranjenimi iz trajnostno usmerjenega sistema, kot je aplikacija TestbedXMaiL. Izračunani stroški, ki zajemajo prevzem in vstavev metapodatkov, znašajo do 2,30 evra za el. sporočilo, stroški za e-sporočilo XML pa znašajo 0,06 evra. Pri tem je pomembno, da ti zneski temeljijo na spremenljivkah, uporabljenih v stroškovnem modelu (npr. serija z 200 e-sporočili ter 4 ure dela za prevzem in valorizacijo te serije).

Glavna razlika je, da so e-sporočila XML že opremljena s primernimi metapodatki in strukturo (kot je opisano v shemi XML).

### **Razvoj načina hrambe**

Razvoj načina hrambe predstavlja glavni delež stroškov pri digitalni hrambi. Spodnji primer temelji na hrambi stotih preglednic na leto (5 serij po 20 preglednic). Tu je bistvena razlika med razvijanjem strategije migracije (33,50 evra na preglednico) ter razvijanjem načina UVC – natančneje razvijanjem programa "dekoder podatkovnega formata" (769,10 evra na kos). Uporaba načina XML je brezplačna (s predpostavko, da je na voljo aplikacija za hrambo v XML s primernimi metapodatki).

Razvoj načina hrambe	Stroški osebjna/na uro v evrih		Obstoječa preglednica (migracija)	Obstoječo preglednico pretv. v XML	Obstoječa preglednica (UVC)	Nova preglednica v XML
Zbiranje zahtev	35,63	stroški za vrsto dokumenta	285,00	285,00	285,00	0,00
Razvoj načina	35,63	stroški za vrsto dokumenta	2850,00	4275,00	71250,00	0,00
Testiranje načina	26,88	stroški za vrsto dokumenta	215,00	322,50	5375,00	0,00
Stroški v evrih za vrsto dokumenta			3.350,00	4.882,50	76.910,00	0,00
Stroški za kos			33,50	48,83	769,10	0,00

Tabela 1: Stroški razvoja načina hrambe

Način hrambe	Obstoječa preglednica (migracija)	Obstoječa preglednica pretv. v XML	Obstoječa preglednica (UVC)	Nova preglednica v XML
Velikost serije	200	200	200	200
Število serij na leto	5	5	5	5
Skupni stroški v evrih na leto	9.213,79	11.252,29	107.061,79	855,00
Skupni stroški v evrih po 20 letih	709.461,96	337.568,75	2.141.253,83	129.622,92

Tabela 2: Primerjava štirih načinov hrambe

Način hrambe	Obstoječe besedilo (migracija)		Obstoječe besedilo, pretv. v XML		Novo besedilo v XML	
Velikost serije	200	2000	200	2000	200	2000
Število serij na leto	10	10	10	10	10	10
Skupni stroški v evrih na leto	20.479,33	112.579,33	24.574,33	116.674,33	1.710,33	1.710,00
Skupni stroški v evrih po 20-tih letih	1.578.294,67	8.669.994,67	737.230,00	3.500.230,00	279.943,33	1.200.943,33

Tabela 3: Trije načini hrambe in učinki velikosti serije

## Aktivnosti digitalne hrambe

Operativni stroški za izvajanje aktivnosti hrambe so minimalni v primerjavi s stroški za razvijanje dejanskih načinov hrambe. Za namene ocenitve so potrebna avtomatska orodja za ocenitev rezultatov posebnih aktivnosti hrambe. Ročna ocenitev se lahko uporablja za redko testiranje serije "shranjenih" dokumentov.

## Zaključki

Migracija (vključno s konverzijo besedilnih dokumentov v PDF) je dolgoročno drag način hrambe, vzrok je dejstvo, da bo treba tekom časa stroške za razvoj obnavljati. V našem modelu predpostavljamo, da bo morala biti migracija ponovljena vsako tretje leto.

Uporaba načina UVC je v našem stroškovnem modelu najdražji način hrambe, to pa povzroča visoke stroške za razvoj (R&D). Pri projektu "Testbed" smo ugotovili, da je za excel zelo težko razviti program "dekoder podatkovnega formata." Poleg

tega mora biti za vsak (nov) datotečni format razvit (nov) program "dekoder podatkovnega formata". Po drugi strani pa se ti stroški tekom časa ne bodo obnavljali: ko bo program "dekoder podatkovnega formata" razvit bo trajal "večno". Mednarodno sodelovanje na tem področju lahko okrepi uvajanje načina UVC.

Na osnovi našega stroškovnega modela lahko trdimo, da se model XML kaže kot cenovno najugodnejši način hrambe, v primerjavi z migracijo (kompatibilnost za nazaj<sup>9</sup> in pretvorba v PDF) ter uporabo načina UVC, še posebej če so dokumenti oblikovani neposredno v XML. Nenazadnje, tudi dokumenti, ki so oblikovani v XML (ali konvertirani v XML), potrebujejo neko vrsto aktivnosti hrambe, ker XML ne bo trajal večno. V našem stroškovnem modelu je predvideno, da bodo dokumenti v XML (oblikovani neposredno v XML ali konvertirani v XML) po 10 letih potrebovali pretvorbo v nov odprt standard.

*Prevod: mag. Natalija Glažar, Arhiv Republike Slovenije  
Strokovna redakcija: mag. Sonja Jager,  
Ministrstvo za javno upravo RS*

---

<sup>9</sup> Ang. Backward compatibility, op. prev.