

PROCES PRENOSA ELEKTRONSKEGA ARHIVSKEGA GRADIVA V NORVEŠKI DRŽAVNI ARHIV

*Terje Pettersen Dahl**

UDK: 930.253:004.3

Terje Pettersen Dahl: Proces prenosa elektronskega arhivskega gradiva v Norveški državni arhiv. Tehnični in vsebinski problemi klasičnega in elektronskega arhiviranja. Zbornik referatov z dopolnilnega izobraževanja, Maribor 4/2005, št. 1, str. 128-134.

Izvirnik v slovenščini, izvleček v slovenščini, povzetek v angleščini.

Prispevek predstavlja proces prenosa elektronskega arhivskega gradiva. Obravnava vse faze od ugotavljanja elektronskega gradiva državne in javne uprave pa do dolgoročne hrambe zapisov v obliki elektronskega gradiva. Prispevek vključuje tudi seznam standardov in orodij, ki se uporabljajo v procesu testiranja.

VREDNOTENJE IN HRAMBA

Prvo fazo procesa vrednotenja predstavlja pregledovanje različnih sistemov, ki so v uporabi. Javne institucije na Norveškem morajo obvezno obvestiti Norveški državni arhiv o svojih sistemih, vendar je za popoln vpogled v stanje potrebno obiskovati te institucije in izvajati nadzor.

Na splošno to pomeni, da je potrebno izvajati vrednotenje obstoječih sistemov, ki običajno nimajo funkcij, povezanih s hrambo gradiva. Zaradi pomanjkanja teh funkcij v sistemu, lahko pride do izgube dragocenih informacij. To onemogoča pridobivanje informacij o poteku zapisov. Če sistem ne vsebuje funkcij zajemanja, mora biti hramba omejena na posnetke stanj dokumentov.

Težko je tudi identificirati logične zapise v tehnični strukturi sistema, in to zaradi pomanjkanja funkcij zajemanja. V nekaterih primerih pa ni na razpolago niti sistemske dokumentacije.

Da bi lahko omogočili ocenjevanje je državni arhiv pričel s projekti pregleda celotnega javnega sektorja (policija, šolstvo, davčna uprava, zdravstvo, kmetijstvo idr.). Ena izmed izkušenj je, da si veliko večino elektronskih sistemov delijo različne institucije znotraj enega sektorja. Posamezni projekti se lahko ukvarjajo samo z elektronskim arhivskim gradivom ali pa vključujejo tako papirno kot elektronsko gradivo. Projekti potekajo v tesnem sodelovanju z javnimi institucijami in njihovimi delavci, ki delajo z dokumentarnim gradivom. Rezultat projektov bo bodisi določitev arhivskega gradiva in njegovega prevzema v državni arhiv, ali pa izločitveni načrt. Namen projekta je identifikacija večuporabniških sistemov s centralno podatkovno zbirko, kombiniranih večuporabniških sistemov z lokalnimi klasičnimi arhivi, istovetnih sistemov v lokalni uporabi z lokalnimi podatkovnimi zbirkami in unikatnih sistemov (lokalnih ali centralnih).

V bodoče nameravamo ocenjevanje sistemov izvajati celo pred začetkom uporabe sistema. Zato bomo morali identificirati informacije (logične zapise), ki jih

* *Terje Pettersen-Dahl, Adviser at The National Archives of Norway, Department of electronic archives, Postbox 4013 Ullevål Stadion, N-0806 Oslo, Norway.*

je potrebno hraniti in jih kasneje predati v državni arhiv. Na ta način bomo lahko zagotovili, da bodo funkcije zajemanja in transferja (predaje) gradiva arhivske vrednosti vključene v sistem.

ZAKONODAJA IN PREDPISI

Na Norveškem velja državni arhivski zakon, po katerem mora vsaka javna institucija predati arhivsko kopijo svojega sistema, ki ga državni arhiv določi kot pomembnega za hrambo.

Arhivski zakon dopolnjuje splošno navodilo o evidentiranju, valorizaciji in predaji gradiva Državnemu arhivu. Ta navodila predpisujejo standard *Noark* kot obvezen za elektronske evidence in za pisarniške sisteme s funkcijami vodenja evidenc.

Dodatno zakon dopolnjuje tudi izredni predpis o predaji elektronskega gradiva. Ta predpis podpira standarde Državnega arhiva za metapodatke in formate. Predpis je v angleški verziji dostopen na www.arkivverket.no/noark-4/Regulation-ch8.pdf. Predpise o gradivu pa lahko najdemo na www.arkivverket.no/noark-4/Regulation-ch9.pdf.

Arhivski zakon definira koncept arhivskega gradiva kot dokument, ustvarjen kot del poslovne aktivnosti, s splošno izjemo dokumentov, ki nimajo poslovne ali dokumentarne vrednosti. Če sistem, postopki, mediji in formati hrambe ustrezajo zahtevam v navodilu, je dokumentarno gradivo lahko elektronsko. Ko sistem preneha delovati, je kopijo elektronskega gradiva potrebno predati državnemu arhivu kot depozit. Če je sistem v uporabi nekaj let, potem je potrebno gradivo predajati v petletnih intervalih. V sistemih odlaganja, ki ustrezajo norveškemu standardu *Noark* je to urejeno tako, da se gradivo razporeja v posamezna petletna obdobja. Celote zaključenega gradiva za posamezna obdobja nato sukcesivno deponiramo.

Glede na arhivski zakon določa državni arhiv/arhivist specifične predpise v zvezi s hrambo ali izločanjem. Predpisi o izločanju določajo, da je potrebno vso gradivo, ki ni omenjeno, hraniti. Ta pravila so uspešna za arhivsko gradivo na papirju, za elektronsko arhivsko gradivo pa samo pravila o izločanju ne zadostujejo. Elektronski sistemi zahtevajo pravila o hrambi, pri tem pa je potrebno več, kot samo golo ime sistema. Predpisi morajo določati logično gradivo, ki ga je potrebno hraniti v primeru, ko gre za sistem za upravljanje z dokumentarnim gradivom. Drugače pa morajo predpisi določati vsaj natančne tipe informacij ali funkcij, ki jih je potrebno hraniti.

V navodilih ni potrebno določiti natančnih metod predaje gradiva, vendar morajo vsebovati navodila za določitev natančnih metod. Predajo bi morale biti mogoče omejiti samo na logično gradivo/zapise (ali druge določene elemente) in potemtakem urediti izločanje. Zaželeno je, da se gradivo arhivske vrednosti v sistemu odlaga sukcesivno med nastankom.

USTVARJANJE IZPISOV ZA PRENOS

Načrti, nastali na podlagi projektov vrednotenja, ki jih je izvedel državni arhiv, določajo trajno gradivo različnih vej uprave. Načrti predpisujejo upravnim institucijam, iz katerih sistemov je potrebno izdelati izpise za prenos v državni arhiv.

Ker je nemogoče prevzeti celoten sistem, izdelamo izpise natančno določenih informacij, ki jih vsebuje sistem.

Institucije morajo izdelati izpise same z lastnim osebjem ali pa si morajo pomagati z najetimi konzultanti. Proces izdelave izpisov lahko zahteva prilagajanje posameznih določil v zvezi z roki hrambe. Izpise nato v obliki arhivske kopije predamo državnemu arhivu.

Izpisi za prenos vsebujejo samo informacije iz tabel podatkovnih zbirk. Oblike, v kateri so v sistemu predstavljene informacije in logika teh oblik se ne ohrani. Informacijo je zaradi tega potrebno ločiti od sistema, ki jo upravlja. To pred prenosom v državni arhiv zahteva reorganizacijo in preformatiranje. Informacije je potrebno izpisati v tekstovni datoteki (flat files). V obliki dokumentacije mora poleg tega datoteke spremljati popis strukture in vsebine (tehnični metapodatki), kakor tudi ostala dokumentacija, npr. sistemska dokumentacija in priročniki za uporabo. Z metapodatki in ostalo sistemsko dokumentacijo lahko kasneje obnovimo izvirno logično strukturo in včitamo datoteke. Pomembno je poudariti, da je potrebno opraviti izpis prenosnih datotek še v času delovanja sistema.

Državni arhiv uporablja za opis strukture in ostalih tehničnih elementov prenosa XML sintakso. Ti tehnični metapodatki v elektronski obliki omogočajo samodokumentiranje hranjenega elektronskega gradiva. To omogoča lažje obnavljanje izvirne (logične) strukture podatkovne zbirke in uporabo arhiviranega elektronskega gradiva.

Za podporo procesa dokumentiranja elektronskega prenosa v upravnih institucijah je državni arhiv razdelil elektronski obrazec (*Arkadukt*) za specifikacijo podatkovnih struktur v osnovnem besedilu. Rezultat pa je XML opis v odgovarjajočem formatu (*ADDMML*).

Današnji splošni sistemi za podatkovne zbirke (Oracle, MS SQL Server idr.) nameravajo ponuditi orodja za avtomatsko dokumentiranje izhodnih podatkov v XML formatu. Tako bo mogoče konvertirati opise iz XML-DTD v *ADDMML*, ali pa jih bo mogoče dokumentirati neposredno v *ADDMML* kot izhod iz sistema podatkovne zbirke.

Zavedati se je potrebno, da morajo biti elektronski dokumenti v prenosu povezani z novo strukturo.

ARHIVSKA KOPIJA

Ko državni arhiv prevzame arhivsko kopijo, jo najprej testira. S tem prvim testom ugotovijo, ali je kopija popolna. Poleg testiranja podatkovnih datotek, preverijo tudi, ali prenos vsebuje potrebno dokumentacijo.

V mnogih primerih so tehnični metapodatki prevzeti na papirju in jih je potrebno konvertirati v elektronski format. Konverzijo iz papirnega v elektronski format izvede osebje državnega arhiva z uporabo *Arkadukta*. To konverzijo je potrebno izvesti pred popolnim testiranjem arhivske kopije. Ko so oboji, tako metapodatki kot podatkovne datoteke v elektronski obliki, lahko izvedemo dejansko testiranje.

Z orodjem *Arkade* testiramo doslednost in pravilnost arhivske kopije, razen v primeru, da izvirni sistem temelji na vpisniku, ki deluje na osnovi *Noarka*. S pomočjo *Arkade* lahko podatkovne datoteke konvertiramo tudi v standardni format, ki ga uporabljamo za dolgoročno hrambo (če je to potrebno). Če se kasneje odločimo za

spremembo formata za dolgoročno hrambo, lahko s pomočjo *Arkade* ponovno izvedemo konverzijo.

Če je izvirni sistem register na osnovi *Noarka*, za testiranje uporabimo *ArkN3* ali *ArkN4*. V tem primeru so metapodatki, ki opisujejo tehnično strukturo standardizirani in objavljeni, in jim ni potrebno predstavljati sestavnega dela transferja.

Po opravljenem testiranju arhivsko kopijo shranimo za dolgoročno hrambo, seveda če je sprejeta, v nasprotnem primeru pa moramo dobiti novo, popravljeno kopijo in ponoviti postopek testiranja.

STANDARDI IN ORODJA

Državni arhiv je za podporo upravljanja z elektronskim arhivskim gradivom določil dva standarda:

- *ADDMML* za opis tehnične strukture in prenosa iz sistema podatkovnih zbirk.
- *Noark*, ki določa funkcionalne zahteve in formate prenosa za sisteme upravljanja z dokumentarnim gradivom, ki temeljijo na delovodnikih.

Kot dodatek k tem standardom, je državni arhiv razvil celo vrsto elektronskih orodij za testiranje arhivskih kopij in za opis na različnih stopnjah procesa.

ADDMML (ARCHIVES DATA DESCRIPTION MARK-UP AND MANIPULATION LANGUAGE)

To je lastni standard Državnega arhiva za popis tehničnih metapodatkov, ki spremljajo prenesene izpise podatkovnih zbirk.

Prva verzija *ADDMML* je bila razvita leta 1998, zadnja nadgradnja pa je bila izvedena v obdobju zima/pomlad 2005. Trenutno so veljavne verzije *ADDMML* 7.3 (2002) in 8.0 (2005).

ADDMML je označevalni jezik, ki temelji na XML, definira pa ga DTD. Sestavljen je iz treh večjih delov - referenc, strukture in procesov. Reference opisujejo kontekst izvirnega sistema (kontekstualni metapodatki), čeprav zelo na kratko. Struktura predstavlja glavni del in vsebuje opis arhivske kopije (tehnični metapodatki). Procesi niso del opisa, vendar nam dajejo možnost določanja specifičnih navodil (procesov), ki jih je za testiranje potrebno izvesti na arhivski kopiji (v *Arkade*). Struktura v strukturnem delu je sestavljena iz petih nivojev: podatkovnega niza, datoteke, tipa dokumenta, tipa polja in kode. Podatkovni niz je celotna arhivska kopija, ki je lahko sestavljena iz različnih datotek, vsaka z različnimi tipi dokumentov itd. Tabela v izvirni podatkovni zbirki običajno ustreza datoteki (relacijska podatkovna baza). V nekaterih starejših tipih podatkovnih zbirk (hierarhičnih itd.) tako nivo datoteke kot nivo tipa dokumenta ustrezata izvirni tabeli.

NOARK (THE NORWEGIAN RECORDS MANAGEMENT STANDARD FOR SYSTEMS WITH REGISTRY FUNCTIONS)

Prva verzija *Noarka* je bila razvita 1984, trenutno pa je veljavna verzija 4.1 (1999/2002). Predvidevamo, da bo verzija 5.0 izdana v letošnjem letu.

Noark standard pokriva področje dokumentacije pri ustvarjalcih kot npr. dokumente, ki so opremljeni z metapodatki o njihovi hrambi. Pisarniško dokumentarno gradivo lahko definiramo kot pisma, zapisniki, interna sporočila, e-pošta itd., medtem ko metapodatki vodenja evidenc pomenijo identifikacijo zapisa/gradiva, čas zajetja zapisa/gradiva, vsebino, kontekst in strukturo informacije itd.

Funkcionalne zahteve *Noarka* lahko uvedemo v samostojne registrske sisteme ali v splošne in specializirane pisarniške sisteme, ki vsebujejo funkcije registriranja.

Iz arhivske perspektive je osnovna ideja *Noarka* zagotoviti, da bodo sistemi za hrambo dokumentarnega gradiva ustrezali določenim standardom funkcionalnosti in metapodatkov za hrambo dokumentarnega gradiva in da bodo ustrezali zahtevam za kratko in dolgoročno varstvo arhivskega gradiva. Z vidika upravljanja z dokumentarnim gradivom pa so primerni sistemi za uporabo, ki temeljijo na priznanih principih hrambe dokumentarnega gradiva, kakor tudi standardizacija v povezavi s spremembami v organizaciji, izobraževanju osebja, vzdrževanju sistema in transferja v skladišča.

Standard *Noark* določa informacijske elemente (katere metapodatke lahko ali mora vključevati sistem *Noark*), osnovne podatkovne strukture (strukturo teh informacijskih elementov in odnose med njimi) in funkcionalnost (katere funkcije mora, vsaj minimalno, podpirati sistem *Noark*). Standard ne določa uporabniškega vmesnika (na kakšen način komunicira sistem z uporabnikom), izvedbe podatkovnih struktur (tehnično oblikovanje), funkcionalnosti (razen specificiranih osnovnih zahtev) ali splošne tehnične izvedbe.

Vsi *Noark* sistemi morajo vsebovati izvozne funkcije za metapodatke in elektronske dokumente o hrambi gradiva. Obstajata dva različna izvozna formata, eden za prenos podatkov iz enega *Noark* sistema v drugega, in drugi, za prenos v skladišče. Za metapodatke o hrambi gradiva je izvozni format določen v XML. Za elektronske dokumente pa mora biti izvozni format eden izmed formatov dokumenta, ki jih za arhivske namene potrди državni arhiv. Trenutno so potrjeni arhivski formati ISO 8859-1 (osnovno besedilo), TIFF, XML in PDF. Vsi *Noark* sistemi morajo vsebovati tudi funkcije za uvoz podatkov iz drugih, z *Noarkom* kompatibilnih sistemov. Uvoženi metapodatki in dokumenti pa so dostopni tako kot v izvornem sistemu.

Naslednja generacija *Noarka* (5.0) pa naj ne bi bila omejena samo na upravljanje z dokumentarnim gradivom, ki temelji na delovodnikih, temveč bi naj služila tudi za razširjanje funkcij upravljanja z dokumentarnim gradivom. *Noark* naj bi uporabljal osnovni model, ki je dovolj splošen in fleksibilen za ta namen. Osnovni model bi moral podpirati različne poslovne procese in upravljanje z evidencami, ter bi naj bil samo eden izmed teh procesov. Osnovni model bi moral podpirati različne procese paralelno. Model bo delal na podlagi ISO 15489, MoReq in norveškega standarda za evidence zdravstvene službe (EPJ).

Naslednja generacija *Noarka* bi morala imeti sposobnost boljšega vzajemnega delovanja z drugimi specializiranimi pisarniški sistemi. Potencial arhiviranja elektronskih dokumentov bi moral biti veliko bolj izrabljen, podpirati pa bi moral različne tipe komunikacije (trenutno veljavna generacija podpira samo elektronsko pošto).

Prvo poglavje standarda *Noark* je dostopno tudi v angleškem jeziku na www.arkivverket.no/english/electronic.html.

ASTA IN ARKIMEDES

Asta je sistem državnega arhiva za popisovanje in izvajanje vseh vrst prenosov. *Asta* v sedanji verziji ne upravlja z elektronskim arhivskim gradivom, zato uporabljamo za tovrstno gradivo *Arkimedes*. Nova verzija *Aste* bo pripravljena v letu 2005 in bo nadomestila tako *Asto* (staro verzijo) kot *Arkimedes*.

Namen nove *Aste* (kot tudi *Arkimedes*) je, da bo vseboval logični opis arhivskega gradiva in upravljanje s skladišči. Logični opis je usklajen z mednarodnim standardom za popisovanje arhivskega gradiva ISAD(g). Sistem upravljanja s skladiščem pa mora upoštevati, da ima elektronsko arhivsko gradivo drugačne karakteristike kot gradivo, ki je bilo ustvarjeno in ga hranimo na klasičnih medijih, kot sta papir in pergament.

ARKADUKT

Gre za sistem za registracijo tehničnih metapodatkov, ki ga je razvil Državni arhiv. *Arkadukt* ustvarja ADDMML datoteke.

V primeru, da tehnični metapodatki, ki se nanašajo na prenos gradiva, niso v elektronski obliki, jih je potrebno konvertirati iz papirja. To napravimo z vpisom metapodatkov v *Arkadukt*. Tako ustvarimo elektronsko metapodatkovno datoteko, ki ustreza ADDMML standardu. Državni arhiv omogoča dostop do tega programa brezplačno. S tem pa zagotavlja državnim in občinskim institucijam možnost, da sami ustvarjajo takšne metapodatkovne elektronske datoteke. V nasprotnem primeru mora to opraviti državni arhiv.

Arkadukt ima možnost branja starih ADDMML datotek in jih konvertirati v trenutno verzijo ADDMML. *Arkadukt* pa lahko izvaja logično testiranje metapodatkov in njihove strukture.

ARKADE

Arkade je sistem za preverjanje prenosa v Državni arhiv, ki temelji na ADDMML.

Arkade je razvit v SAS in lahko izvaja različna testiranja in analize arhivskih kopij. Nadzira konsistenco med dejanskimi podatkovnimi datotekami in metapodatki, torej da metapodatki odražajo in dokumentirajo pravilno strukturo podatkovnih datotek in da je pravilna tudi sama informacija o metapodatkih, npr. število dokumentov *Arkade* lahko izvaja analize, preglede in nadzor na različnih ravneh. Dodatno pa lahko izvaja tudi nekaj specialnih funkcij. *Arkade* lahko izvede konverzijo v arhivsko kopijo iz enega niza v drugega (to je, da konvertira niz karakterjev idr.), da ustvari ustreznost verzije za dolgoročno hrambo. *Arkade* lahko ustvari tudi osnovo (SAS podatkovni niz) za javno verzijo arhivske kopije.

ARKN3 IN ARKN4

To so sistemi za testiranje in predstavljanje prenosa, ki temelji na *Noarku* (verzije 3.0 in 4.0).

ArkN3 in ArkN4 uvažata arhivske kopije iz sistema *Noark* in izvajata pregled doslednosti podatkov. Tudi ta dva programa lahko izvajata enostavne analize.

ArkN3 in ArkN4 lahko predstavita informacije tudi za javno uporabo.

SUMMARY

THE PROCESS OF TRANSFERRING ELECTRONIC ARCHIVES TO THE NATIONAL ARCHIVES OF NORWAY

This paper deals with the process of transferring electronic archives from survey to long-term preservation. It covers the main events in the process as survey, creating a plan for preservation (or disposal), transfer, test of archival copy and long-term preservation, and presents how these tasks are handled in Norway. The paper also describes the impact of archival law and regulations, on the process. A description of Norwegian standards and tools used in the process is included.