

OD INFORMACIJSKO TEHNOLOŠKIH SISTEMOV DO ARHIVOV

*Jan Dalsten Sørensen**

UDK: 004.3:930.253

Jan Dalsten Sørensen: Od informacijsko tehnoloških sistemov do arhivov. Tehnični in vsebinski problemi klasičnega in elektronskega arhiviranja. Zbornik referatov z dopolnilnega izobraževanja, Maribor 4/2005, št. 1, str. 115-120.

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

Članek obravnava izzive, ki jih predstavlja digitalizacija državne uprave in opisuje ravnanje Danskega državnega arhiva pri valorizaciji in prevzemanju elektronskega gradiva. Avtor izpostavlja nekatera področja razvoja in prioritete.

UVOD¹

Razvoj v smeri vedno bolj razširjene e-uprave nas postavlja pred vprašanja zagotavljanja pravilne identifikacije elektronskih dokumentov, njihove predaje v arhiv, hrambe, dostopnosti in berljivosti tudi v prihodnosti. Kakšne spremembe v arhivski praksi so potrebne, da bomo lahko izpolnili zahteve, ki jih postavljajo nove okoliščine? Kako bomo obravnavali različne probleme, npr. tehnološko zastarelost? Kako bomo prešli od začetnega spoznavanja izzivov do aktivnega ravnanja z elektronskim gradivom v takšni obliki, da bomo obvarovali informacije, ki so vredne, da jih ohranimo, za prihodnje generacije.

Prvi prevzem elektronskega gradiva je bil v Danskem državnem arhivu izveden sredi sedemdesetih, kar pomeni, da ima arhiv že dolgo tradicijo postavljanja zahtev državni upravi v zvezi s strukturo in uporabo sistemov za upravljanje z dokumenti. Leta 1995 so kot prvi arhiv pripravili specifikacijo za predajo elektronskih sistemov za upravljanje z dokumenti, tako da so tisti uradi, ki so to želeli, lahko zavrgli papirno gradivo.

IDENTIFIKACIJA, OCENA IN ODOBRITEV

Ugotovili smo, da je potrebno informacijsko-tehnološke sisteme, ki jih uporablja uprava, razdeliti v tri osnovne kategorije:

- Podatkovne zbirke: te lahko vsebujejo kakršnekoli informacije, ki jih je zbral urad pri svojem delu, vendar se ne uporabljajo za namene upravljanja z dokumenti.
- Tradicionalni sistemi upravljanja z dokumenti (DMS), ki digitalno hranijo metapodatke o dokumentih in zadevah; dejanski dokumenti pa so fizično na papirni podlagi in zaradi tega niso del elektronskega sistema.

* Jan Dalsten Sørensen, M.A., Special consultant, Head of Section of Electronic Records, Department of Appraisal and Transfer, Danish National Archives, Rigsdagsgården 9, DK - 1218 København K, Denmark.

¹ Pričujoči prispevek je razširjena verzija mojega referata "Javni predpisi o elektronskem gradivu: predpisi in zahteve za javno upravo na Danskem", ki sem ga predstavil na Mednarodnem arhivskem kongresu na Dunaju leta 2004.

- Elektronski sistemi za upravljanje z dokumenti (EDMS), kjer so tako dokument kot tudi relevantni metapodatki shranjeni v digitalni obliki.

Elektronsko gradivo definiramo v najširšem smislu kot digitalne informacije, shranjene v enem izmed zgoraj omenjenih sistemov. Poudariti je potrebno, da danska arhivska zakonodaja z besedo »record« (dokument ali zapis) opisuje vse informacije, ki jih katerikoli urad pri svojem delu zbere in shrani v strukturirani obliki.

Če govorimo o papirnatem gradivu, ga arhivisti običajno ne vidimo prej, preden ni staro vsaj dvajset ali trideset let, včasih celo še veliko kasneje. In tako je tudi prav. Očitno je, da imajo posamezni uradi postopke, ki zagotavljajo, da so dokumenti izdelani in shranjeni na pravilen način, tako logično kot fizično. Običajno arhivistom ni potrebno skrbeti za papirno gradivo dokler je v aktivni rabi oziroma prvih deset ali dvajset let po tem.

Pri elektronskem gradivu je zgodba drugačna. Predstavljalo bi velik izziv, če bi želeli podatke, shranjene v informacijsko-tehnoloških sistemih hraniti dvajset do trideset let, jih nato migrirati na veljavne standarde in predati arhivu v trajno hrambo. Pri tem obstaja veliko tveganje tehnološke zastarelosti, kot so zastareli formati, izguba podatkov zaradi uporabe nepodprte strojne opreme in cela vrsta drugih situacij, ki ogrožajo podatke.

Prav tako je potrebno upoštevati evropsko direktivo o varovanju podatkov. Določila direktive o varovanju podatkov - vsaj kolikor so vključena v dansko zakonodajo - pomenijo, da so lahko osebni podatki shranjeni v sistemu posameznega urada le tako dolgo, kot je to nujno potrebno. V praksi to pomeni, da je potrebno omenjene podatke izbrisati najkasneje v petih letih po zadnjem stiku z obravnavano osebo. Vendar imajo te informacije pogosto zgodovinsko vrednost, kar določa oz. priznava tudi danski zakon o zaščiti podatkov. Rešitev je, da je takšne podatke potrebno prevzeti v arhiv pred izbrisom. To sili arhive v zgodnje prevzeme podatkov, kar pomeni, da je potrebno čim prej identificirati tovrstne dokumente. Zato morajo danske državne službe pred začetkom uporabe novega sistema, o svoji nameri obvestiti Danski državni arhiv.

Če je novi sistem podatkovna zbirke brez vsakih funkcij upravljanja z dokumenti, je dovolj, da urad obvesti državni arhiv dva tedna pred začetkom uporabe. Obvestilo mora biti oddano v obliki posebnega obrazca. Vsebovati mora informacije o namenu sistema in opis podatkov v njem, pa tudi druge informacije, ki jih potrebujemo za proces valorizacije.

Za podatkovne zbirke ne potrebujejo posebnega dovoljenja državnega arhiva, vendar morajo poslati obvestilo zato, da lahko čim prej izvedejo valorizacijo. Če ugotovijo, da so podatki tako pomembni, da jih bodo ohranili, se datum za prevzem podatkov iz sistema določi pri valorizaciji. Običajno prvi prevzem sledi pet let po postopku. Arhiv obvesti urad o tehničnih specifikacijah za prevzem, tako da imajo na razpolago dovolj časa, da ugotovijo, ali obstajajo kakšne težave v zvezi s konverzijo podatkov iz njihove podatkovne zbirke v arhivsko verzijo.

Če gre za nov sistem za upravljanje z dokumenti, mora biti ocenjen in odobren. Zaradi tega je potrebno državni arhiv obvestiti o uvedbi novega sistema za upravljanje z dokumenti najmanj tri mesece pred predvidenim pričetkom uporabe.

Obvestilo mora biti oddano v obliki obrazca, vsebovati pa mora:

- opis sistema za poizvedovanje (npr. klasifikacijski načrt, indeksne označbe, metapodatki, potrebni za poizvedovanje dokumentov),
- osnutek navodil za uporabo sistema,
- tehnično dokumentacijo.

Če sistem ni uveden kot relacijska podatkovna zbirke, mora urad že ob potrditvi oddati arhivu opis, kako je mogoče sistem konvertirati v relacijsko podatkovno zbirko vsaj v prvi normalni obliki, saj bo gradivo navsezadnje na ta način predano v arhiv.

Sistem mora biti razdeljen v arhivska obdobja približno petih let. Ko arhivska doba poteče, morajo biti zaprte vse zadeve, podatkovna zbirke z metapodatki in če je primerno, tudi digitalni dokumenti. Arhivsko kopijo je nato potrebno predati državnemu arhivu. Če so zadeve na papirju, morajo biti vsi dokumenti tega obdobja odloženi kot povezana in nepretrgana serija, za novo arhivsko obdobje pa morajo biti ustvarjeni novi združeni dokumenti (files).

Uvedba sistema mora biti dobro definirana, npr. ali bodo sistem uporabljali vsi uradi ali samo nekateri. Če del posamezne organizacije uporablja tradicionalni sistem upravljanja z dokumenti, kjer so posamezne zadeve na papirju, medtem ko drugi del iste organizacije uporablja popolnoma elektronski sistem upravljanja z dokumenti, kjer so vsi dokumenti v digitalni obliki, je natančno definiranje bistvenega pomena.

Četudi je v okviru organizacije vsesplošno v uporabi popolnoma digitaliziran sistem upravljanja z dokumenti, je potrebno, da omogoča vpisovanje dokumentov, najdenih na papirju (npr. dokumenti nenavadnih oblik, ki jih ni mogoče skenirati v sistem ali poročila, pri katerih skeniramo le naslovnico in kazalo). Zato morajo imeti vsi sistemi posebno polje, v katero je mogoče vpisati, ali je dokument v celoti shranjen v digitalni obliki, delno digitaliziran ali na papirju.

Sistem mora biti narejen tako, da zagotavlja varovanje dokumentov, ki so shranjeni v sistemu. To pomeni npr. da moramo natančno vedeti, katere formate dokumentov lahko shranimo (Word, PDF, TIFF, Excel tabele itd.). V sistemu morajo obstajati samo takšni dokumenti, ki jih lahko konvertiramo v TIFF. Če o tem dvomimo, mora urad zagotoviti pisno izjavo upravljalca programske opreme, da je možno hraniti le tiste dokumente, za katere upravljalec zagotavlja konverzijo. Predlagamo, da se konverzija v TIFF izvede čim prej po registraciji dokumenta. Vse naštetu mora biti vključeno v navodilo o uporabi, tako da so vsi uporabniki seznanjeni s tem, kakšne formate je mogoče shraniti v sistem, če bodo konvertirani v TIFF.

Tudi varnostni ukrepi, ki jih morajo uradi sprejeti v zvezi z izgubo ali ponarejanjem informacij, morajo biti opisani v navodilih. Lep primer so opombe, ki se uporabljajo v wordovih dokumentih: pri konverziji v TIFF se bodo informacije izgubile ali pa bodo opombe zakrile tekst v dokumentu. Možna rešitev je, da shranimo informacije pri konverziji dokumenta v TIFF kot končne opombe.

Prav tako pa lahko zahtevamo od urada, da nam opišejo, kako bodo zavarovali dokumente v primeru, da se njihove aplikacije, npr. wordova programska oprema, v posameznem arhivskem obdobju spremenijo. Pri tem bi bila dobra rešitev, da se pred uporabo nove programske opreme vsi dokumenti konvertirajo v TIFF.

Sistem za poizvedovanje mora biti prav tako opisan v navodilih in ga je potrebno zagotoviti zato, da je možno dokumente, ki sodijo skupaj, tako tudi najti. Navodilo mora opisati metapodatke, ki jih je potrebno najti za dokumente in datoteke.

Poudariti je potrebno, da upravljalec programske opreme ne more dobiti odobritve za posamezen sistem za upravljanje z dokumenti kot tak. Odobritev se daje za uvedbo sistema, pri tem pa se kot celota upoštevajo tehnične rešitve, vkjučno z navodili, klasifikacijskim načrtom itd. Dobro delujoč sistem bi lahko bil uveden na zelo neprimeren način, zato ni dovolj, da gledamo samo na sistem kot tak.

Zahteve za odobritev novih sistemov se nanašajo samo na državne ustanove. Vendar pa je tudi lokalna uprava na podlagi arhivskega zakona dolžna zagotavljati varstvo elektronskega gradiva tako, da ga je možno prevzeti v javne arhive. Od maja 2004 veljajo tehnične specifikacije in zahteve za prevzem vseh javnih podatkov v državni arhiv, ne glede na to, ali gre za gradivo državne ali lokalne uprave in ne glede na to, kateri arhiv bo prevzel gradivo.

PREDAJA

Ko gradivo prevzame državni arhiv, ga je potrebno migrirati na od sistema neodvisen format, ki temelji izključno na standardih. Glavni poudarki so sledeči:

- podatki morajo biti prevzeti kot od sistema neodvisna arhivska verzija,
- podatki morajo biti strukturirani kor relacijska podatkovna zbirke,
- obstajati mora dovolj metapodatkov za opis strukture in vsebine,
- priložene morajo biti splošne informacije o sistemu,
- dokumenti morajo biti konvertirani v standardiziran format,
- če gre za sistem za upravljanje z dokumenti, je potrebno vključiti nekatere SQL-poizvedbe,
- in za konec je potrebno podatke prenesti na CD-R ali na prenosni USB ključ.

Podatki iz vsake tabele morajo biti konvertirani v standardni niz znakov (ISO 8859 - Latin 1) in shranjeni kot sekvenčne datoteke.

Posledica migracije na od sistema neodvisno arhivsko verzijo je izguba funkcionalnosti aplikacije programske opreme, s katero so bili ustvarjeni podatki. Zato je pomembno, da se podatki prenesejo iz sistema z ustreznimi metapodatki in kontekstualnimi informacijami. Med najpomembnejšimi deli metapodatkov je strojno berljiva XML-datoteka z opisom vseh tabel in polj, ki temeljijo na EBNF (Extended Bachus-Naur Form). Na ta način zagotovimo, da so tabele vseh arhivskih verzij opisane popolnoma enako, ne glede na to, s katerim sistemom smo jih generirali. To je seveda potrebno, če želimo napraviti neke vrste standardizirano orodje za poizvedovanje za vse arhivske podatke.

Opis tabel z metapodatki pa ni dovolj, da bi zagotavljal nadaljnjo uporabo podatkov. Zato je potrebno zagotoviti dodatno dokumentacijo in vključiti izbor splošnih informacij v arhivsko verzijo. Za to ni bilo mogoče izdelati enotnega standarda, kajti podatki prihajajo iz različnih sistemov, kjer so tipi razpoložljive in potrebne dokumentacije enako različni. Ne glede na to, poskušamo zagotoviti, da splošne informacije vsebujejo administrativni opis sistema kakor tudi tehnično dokumentacijo, ki ilustrira strukturo in funkcionalnost sistema, npr. ER-diagrami in zajete zaslonske slike (screen dumps). Vse te informacije morajo biti shranjene v TIFF datotekah.

Specifične arhivske informacije najdemo tudi v obliki opisnih tabel s podatki o npr. provenienci, časovnem obdobju, v katerem so bili podatki generirani in ime sistema. Tukaj je tudi tabela, v kateri je indeks dokumentov, ki zagotavljajo splošne informacije.

Kot smo omenili že prej, je potrebno vse dokumente konvertirati v TIFF. Dokumenti morajo biti zgoščeni. Če sistem vsebuje tudi zvočne in video zapise, jih je potrebno arhivirati kot MP3 za audio in MPEG-3 za video zapise.

Če gre za sistem upravljanja z dokumenti, je v metapodatke potrebno vključiti tri do deset najpogostejših vprašanj, da omogočimo bodočim uporabnikom, da najdejo pravilno informacijo, tudi ko izvirna aplikacija ni več na razpolago.

Ko je gradivo predano, ga je potrebno najprej testirati in ugotoviti, ali so bile izpolnjene vse zahteve. Uradi iz svojega informacijsko-tehnološkega sistema ne smejo izbrisati podatkov, dokler državni arhiv ne opravi testiranja in izda potrdila o odobritvi predaje, s čimer prevzame odgovornost za nadaljnjo hrambo gradiva. Če je arhivska verzija pravilna, arhiv izdela dva niza CD-R in jih shrani na dveh različnih lokacijah. Dodatno kopijo arhivske verzije pa shranijo na RAID diske (Redundant Array of Independent Disks).

PODROČJA RAZVOJA IN PERSPEKTIVE

Z arhivskega vidika migracija podatkov iz izvirnega sistema, v katerem so bili ustvarjeni, v arhivsko verzijo, neodvisno od sistema, ne predstavlja idealne rešitve. Glede na gradivo na papirju in posamezne zadeve, želimo arhivisti zagotoviti, da se informacije vzdržujejo in hranijo v avtentični obliki, torej na način, kot so bile ustvarjene in po možnosti na isti način (primerjaj načelo provenience). Migracija podatkov iz izvirnega sistema na arhivsko verzijo, neodvisno od sistema, vključuje spremembo, ki lahko ogrozi avtentičnost dokumenta. Ne glede na to, da ima migracija v od sistema neodvisen format tudi slabe lastnosti, pa moramo biti realni in sprejeti dejstvo, da je to po vsej verjetnosti edina rešitev, ki za sedaj deluje.

Dokler nimamo resnične rešitve za strategijo, ki bi zagotavljala tudi hrambo samih atributov sistema, moramo vedno znova izboljševati in izpopolnjevati splošne informacije, ki so vključene v arhivsko verzijo, da s tem nadoknadimo pomanjkljivosti migracije na arhivsko verzijo, neodvisno od sistema.

Drugo področje, na katerega se je potrebno osredotočiti, pa je razvoj standardiziranega orodja za poizvedovanja, ki bo omogočal boljšo dostopnost do podatkov, shranjenih v arhivski verziji. Seveda je mogoče predelati tabele, recimo v SQL-server, vendar tega ni mogoče storiti brez ustrezne kvalifikacije. Dvomim, da bo mogoče uporabljati podatke iz arhivske verzije brez, vsaj delne, usposobljenosti in prakse, tako kot ni mogoče brez usposobljenosti in tudi mnogo prakse brati rokopisov iz 17. stoletja. Zaradi tega raste potreba po izdelavi orodja, ki bo potencialnim uporabnikom olajšalo ponovno vzpostavitev podatkovne zbirke in povpraševanje brez dodatnega programiranja. Načrtujemo, da bomo takšno orodje razvili čez nekaj let.

SUMMARY

FROM IT-SYSTEMS TO THE ARCHIVES

The Danish National Archives has received transfers of e-records since the mid-1970s. The article argues, based on the experiences of the Danish National Archives, that it is necessary to

- identify and appraise e-records at a much earlier stage than paper records
- make demands on the systems of the authorities so that the authorities take precautions to keep their e-records in a way so they can be transferred to the archives
- transfer the e-records to the archives at an early stage
- migrate the data to a system independent “archival version” based on standard formats
- continue the development of standards, procedures and tools in order to improve the quality of and access to the archival versions of e-records that are transferred to the archives.