



Titel: Komplet teknisk instruktion för digital åtkomst och digitala leveranser till TAM-Arkiv

Datum: 2011-09-01

TAM 7:2011 INSTRUKTION

Komplett teknisk instruktion för digital åtkomst och digitala leveranser till TAM-Arkiv

Revision 1.0

Innehållsförteckning

1 Kort beskrivning ("Executive summary")	3
2 Inledning	4
3 Kontakt	6
4 Bakgrund	7
5 Åtkomst	8
5.1 Övergripande beskrivning för åtkomst	8
5.2 Detaljerade villkor för åtkomst	9
5.2.1. Åtkomstprocedur	9
5.2.2. Juridiska villkor	9
5.2.3. Finansiering för åtkomst	9
6 Leverans	10
6.1 Övergripande beskrivning för leverans	10
6.2 Detaljerade villkor för leverans	13
6.2.1. Bestäm vad som ska levereras	13
6.2.2. Villkor för framtida åtkomst och söklistor	13
6.2.3. Leveransprocedur	14
6.2.4. Juridiska villkor	14
6.2.5. Finansiering vid leverans	14
6.2.6. Leveransdokumentation	15
6.2.7. Leveransmedia	15
6.2.8. Leveranspaketstruktur	16
6.2.9. Filformat och teckentabeller	17
6.2.10. Databaser inklusive register	17
6.2.11. Kontorsdokument, fotografier, tidskrifter, CAD-filer och liknande	20
6.2.12. E-post	21
6.2.13. Diarium och register	22
6.2.14. Dokumenthanterings- och ärendehanteringssystem	22
6.2.15. Webbsiter, bloggar, sociala media och källkod	23
6.3 Revisionshistorik	25
6.4 Länkar	26
6.5 TAM-Arkivs rekommendationer och rutiner	27

1 Kort beskrivning ("Executive summary")

Denna instruktion beskriver hur digitala leveranser till TAM-Arkiv och åtkomst till digitalt lagrad information vid TAM-Arkiv går till. Denna instruktion är komplett med detaljerade tekniska anvisningar.

TAM-Arkiv tar i organiserad form emot mindre mängder digitalt lagrad information från sina medlemsorganisationer sedan år 2011. Åtkomst till tidigare digitalt inkommen information ges enligt gällande regler för att ta del av arkivhandlingar vid TAM-Arkiv.

TAM-Arkiv har tyvärr ingen möjlighet att ta ansvar för inkommen digital information, om inte TAM-Arkivs instruktioner för digitala leveranser följs. Det är mycket viktigt att en förberedande kontakt tas i god tid innan en digital leverans sker. Detta för att TAM-Arkiv ska kunna hjälpa medlemsorganisationerna att bevara sin information. Sker det inte kan oftast inte inkommen information förvaltas, utan läggs i en arkivbox eller på en hylla i arkivdepån. Detta innebär att inkommen information i detta fall endast är tillgänglig så länge som ursprunglig mediatyp och ursprungligt filformat är läsbart. Detta kan i praktiken ofta innebära så kort tid som några år.

TAM-Arkiv är en ideell förening som ägs av sina medlemsorganisationer. Tyvärr har inte TAM-Arkiv i dagsläget varken någon teknisk plattform eller finansiella resurser att hantera stora, komplexa eller automatiserade digitala leveranser.

2 Inledning

Denna instruktion beskriver hur åtkomst och leveranser av digitalt material till TAM-Arkiv sker. Dokumentet riktar sig till systemarkitekter, arkivarier, IT-arkivarier, informationsförvaltare, informationsägare, systemutvecklare, systemförvaltare och liknande vid TAM-Arkivs medlemsorganisationer. Detta är den kompletta instruktionen för leveranser och åtkomst till digital information vid TAM-Arkiv. För att kunna förstå hela dokumentet förutsätts läsaren ha kunskaper om databasmodellering, XML och IT-system i allmänhet. Har man inte dessa bakgrundskunskaper inom IT kan man i alla fall förhoppningsvis läsa och förstå de avsnitt som beskriver åtkomst och leverans på ett mer övergripande sätt utan tekniska detaljer, se avsnitt 5.1 och 6.1.

TAM-Arkiv tar i organiserad form emot mindre mängder digitalt lagrad information från sina medlemsorganisationer sedan år 2011. Åtkomst till tidigare digitalt inkommen information ges enligt gällande regelverk för att ta del av arkivhandlingar vid TAM-Arkiv. Notera att öppenhet är huvudregeln för levererat material om inte annat specificerats tydligt vid leverans, eller att lagkrav begränsar tillgänglighet.

TAM-Arkiv har tyvärr ingen möjlighet att ta ansvar för inkommen digital information, om inte TAM-Arkivs instruktioner för digitala leveranser följts. Det är mycket viktigt att en förberedande kontakt tas i god tid innan en digital leverans sker. Detta för att TAM-Arkiv ska kunna hjälpa medlemsorganisationen att bevara sin information. Sker inte detta kan ofta inte inkommen information förvaltas, utan läggs i en arkivbox i arkivdepån. Detta innebär att inkommen information i detta fall endast är tillgänglig så länge som ursprunglig mediatyp och ursprungligt filformat är läsbart. Detta kan i praktiken innebära så kort tid som några år.

Stora, komplexa eller automatiserade digitala leveranser har TAM-Arkiv i dagsläget inte tekniska resurser att hantera.

I sitt arbete med digital arkivering följer TAM-Arkiv ISO 14721 "Space data and information transfer systems - Open archival information system - Reference model" och Paradigm-rekommendationerna från universiteten i Oxford och Manchester. TAM-Arkiv använder också delar av de krav som ges av riksarkiven i Danmark, Norge och Sverige, exempelvis delar av danska riksarkivets "Vejledning til bekendtgørelse nr. 1007 af 20. august 2010". TAM-Arkiv använder också olika IT-verktyg som tagits fram eller används vid riksarkiven i Schweiz, Sverige och Danmark. Exempelvis kan man från både danska¹ och svenska riksarkivet² ladda ner verktyg för konvertering och olika former av kontroller. Dessutom

¹ Se "Tea3 og hjælpeprogrammer til produktion af arkiveringsversioner " på sidan http://www.sa.dk/content/dk/for_statslige_myndigheder/aflevering/it-systemer/aflevering_efter_bek_nr_342/tea3_og_hjalpeprogrammer och SABA.

² Se RALF och "Hjälpmiddel för leveranskontroll och arkivuttag vid leveranser av databaser" på Riksarkivets hemsida.

använder TAM-Arkiv andra kommersiella och "open source"-verktyg för att hantera digitala leveranser.

TAM-Arkiv tar endast emot digitalt lagrad information som ska bevaras. Gallringsbart material (det vill säga material som ska förstöras efter en viss tid) tas inte emot annat än i undantagsfall, och efter uttrycklig överenskommelse. Detta eftersom utformande i efterhand av automatiska gallringsregler i många fall kan bli både komplext och dyrbart för TAM-Arkiv, och därigenom i slutändan även för dess medlemsorganisationer. Exekverbar programkod och script får inte levereras till TAM-Arkiv, på grund av att TAM-Arkiv inte kan garantera ursprunglig funktionalitet över tiden, och för att minska risken för spridning av skadlig kod ("malware").

För övergripande information om krav på arkivfunktionalitet i IT-system hänvisas till rekommendationen "TAM4:2010-Arkivfunktionalitet för IT-system". För övergripande information om informationsförvaltning se rekommendationen "TAM2:2008-Informationsförvaltning". Alla rekommendationer kan laddas ner från TAM-Arkivs hemsida, www.tam-arkiv.se.

Ett stort tack går till Monika Jansson vid Riksantikvarieämbetet, Karin Bredenberg, Mattias Djupdahl, Johan Ekdahl, Jens Kling, Christina Olsson och Eva Toller vid Riksarkivet, och Bo Johansson vid Sollentuna kommun, vilka alla på olika sätt har visat på fel och otydligheter i tidigare versioner. Eventuella kvarvarande sakfel står undertecknad för.

Förutom att TAM-Arkiv använder denna instruktion i sina kontakter med medlemsförbunden, så hoppas vi även att andra organisationer som kämpar med den praktiska verkligheten inom området digital arkivering kommer att ha nytta av den som bakgrundsmaterial.

Magnus Wählberg
Stockholm 2011

3 Kontakt

Kontakt för åtkomst till digitalt levererad information och framtida digitala leveranser enligt nedan. Var vänlig att ta kontakta i god tid innan digital leverans, eftersom digitala leveranser kräver mer förberedelsearbete än leveranser på papper.

Email

magnus.wahlberg@tam-arkiv.se

Telnr

08-5454 1562 (Magnus Wåhlberg)

Postadress

TAM-Arkiv
Digitalt
Grindstuvägen 48-50
167 33 Bromma

Hemsida

www.tam-arkiv.se

Kommentarer, påpekanden om otydligheter och felaktigheter i denna instruktion tas gärna emot.

4 Bakgrund

Syftet med långtidslagring är att kunna återsöka, läsa och återanvända digital information på längre sikt, säg 5-1000 år. Digital information, eller samhällets minne om man så vill, ska vara tillgänglig för framtida generationer och samhälle. Om man inte bevarar information blir alternativet ”Digital Darknes”. Termen Digital Darknes innebär att digital information förstörs efter godtycklig tid, och att om något blir kvar i framtiden är endast till följd av rena tillfälligheter.

Det finns en internationell standard för hur arkivsystem och interaktion med arkivsystem ska se ut, den så kallade OAIS-modellen.³ I denna standard beskrivs bland annat hur tillgång till ett arkiv och leverans till ett arkiv ska gå till, se figur 1.

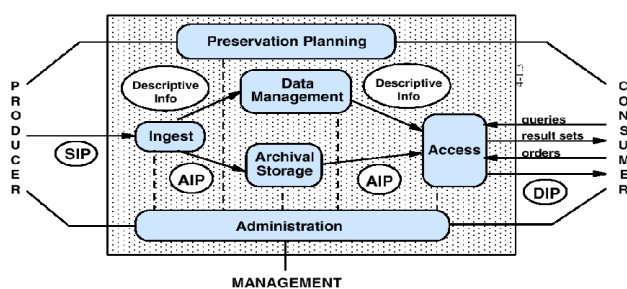


Fig 1. Förhållandet mellan producent, konsument och arkivsystem enligt OAIS-modellen.

TAM-Arkiv följer denna standard bland annat genom att dela upp i åtkomst och leverans. Enligt denna modell levererar en producent, d.v.s. ett medlemsförbund information till TAM-Arkiv, där informationen sedan förvaltas enligt OAIS-modellen. När en konsument, d.v.s. en forskare eller ett medlemsförbund, sedan önskar få tillgång till informationen sker även detta enligt denna modells beskrivningar.

3 ISO 14721:2003, Space data and information transfer systems -- Open archival information system -- Reference model.

5 Åtkomst

5.1 Övergripande beskrivning för åtkomst

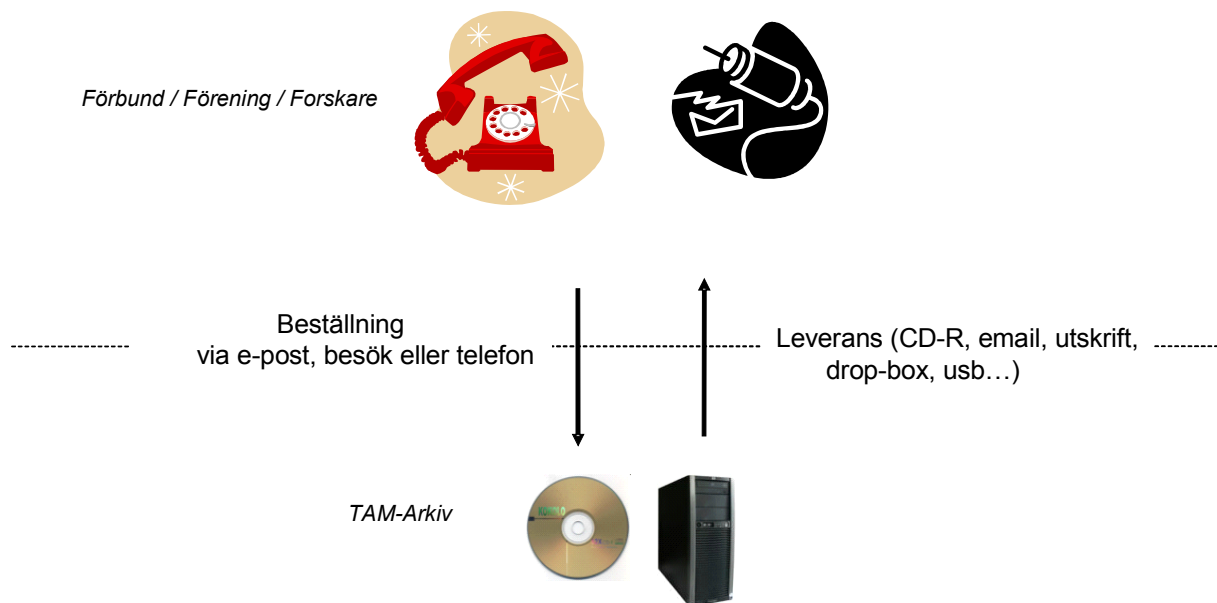


Fig 2. Schematisk beskrivning för åtkomst till digitalt material vid TAM-Arkiv.

Arbetssteg som görs innan åtkomst:

1. Maila, besök eller ring ansvarig för digital arkivering vid TAM-Arkiv. Se kapitel "Kontakt".
2. Klargör vad som önskas beställas, önskat format och önskat leveranssätt.
3. Teckna en enkel överenskommelse om digital åtkomst, med överenskomna villkor för leveransen/-erna från TAM-Arkiv, om detta inte finns sedan tidigare.

Arbetssteg som görs vid TAM-Arkiv:

- a) Bedöm om informationen kan lämnas ut. Behövs tillstånd från medlemsorganisation? Kan det finnas uppgifter som inte bör lämnas ut med hänsyn till enskild?
- b) Eventuell omarbetning från arkivformat till leveransformat.
- c) Paketera digitalt eller skriv ut på papper (endast för mindre mängder).
- d) Leverans sker på överenskommet sätt.
- e) Uppdatering av åtkomststatistik.

Kompletterande villkor för digital åtkomst ges i följande avsnitt "Detaljerade villkor för åtkomst" i denna instruktion.

5.2 Detaljerade villkor för åtkomst

5.2.1. Åtkomstprocedur

Bedöm och dokumentera i överenskommelse:

- Bedöm hur stora återsökningsbehoven är. Hur ofta och hur snabbt behöver man tillgång till informationen.
- När ska åtkomst ske? Dygnet runt?
- På vilket sätt? Via telefon, email, besök?
- Vem får åtkomst? Behövs behörighetskontroll?
- Vilka tjänster önskas tillgång till? (Queries, orders, reports...)

Informationen från detta avsnitt dokumenteras i överenskommelsen mellan part som önskar få tillgång till digitalt arkivmaterial och TAM-Arkiv.

5.2.2. Juridiska villkor

Generellt är det öppenhet som gäller vid åtkomst till tidigare levererad digital information. Det vill säga fri tillgång ges till tidigare information. Öppenhet gäller dock inte om ett förbund vid leverans begärt att vissa serier inte ska lämnas ut. Öppenhet gäller inte heller om TAM-Arkiv bedömer att ett utlämnande kan stå i strid med svensk lagstiftning eller om det finns risk att en enskild kan lida skada. Exempelvis så lämnas inte arbetsskadeakter, enskilda löneuppgifter och liknande personrelaterad information ut till forskare.

Vad gäller användande av upphovsrättsskyddat material, såsom vissa fotografier och filmer, så måste mottagande part kontrollera med den som har upphovsrätten till materialet under vilka villkor som det får användas. I en del fall vet TAM-Arkiv om ett material (verk) omfattas av aktuell copyright och upplyser då om detta till mottagande part. I många fall är det tyvärr oklart om ett material omfattas av någon aktuell form av copyright och vem som då har copyright. TAM-Arkiv hoppas att få resurser och möjlighet att utreda denna komplexa fråga i framtiden.

5.2.3. Finansiering för åtkomst

Eventuella extrakostnader för åtkomst debiteras enligt gällande prislista och betalningsmodell.

Informationen från detta avsnitt dokumenteras i överenskommelsen mellan part som önskar få tillgång till digitalt arkivmaterial och TAM-Arkiv.

6 Leverans

6.1 Övergripande beskrivning för leverans

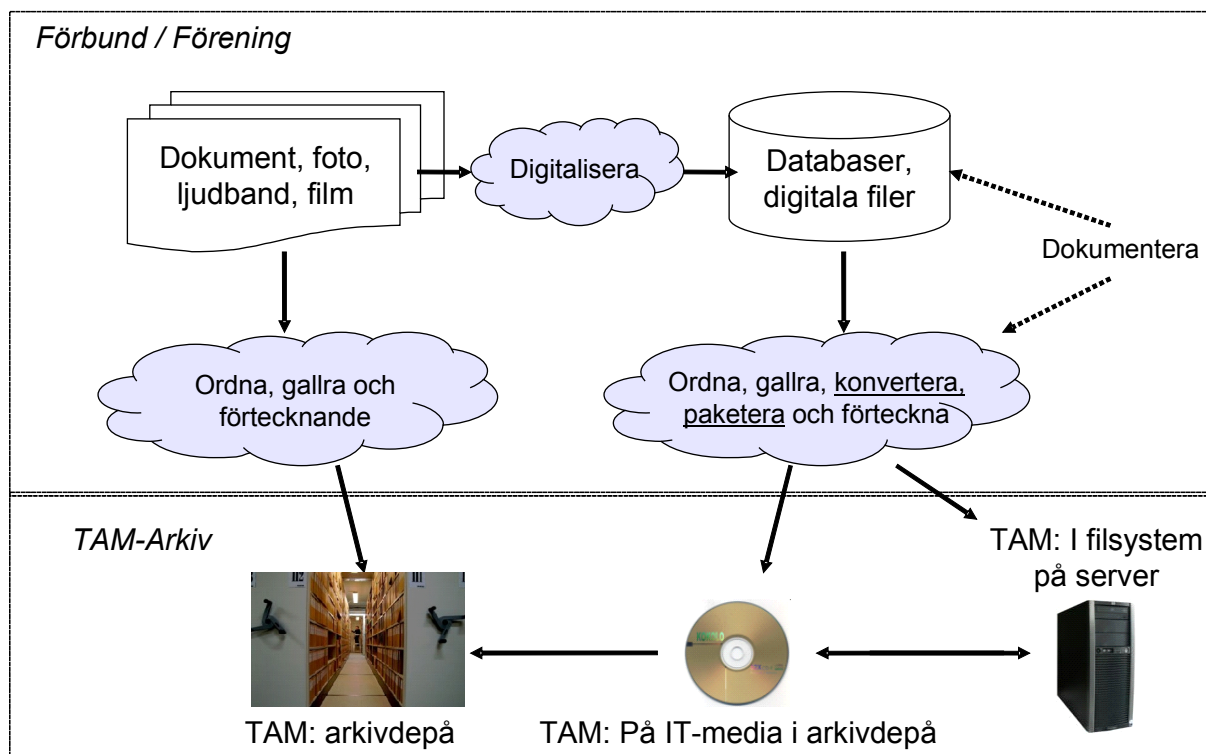


Fig 3. Schematisk beskrivning för leverans av analogt och digitalt material vid TAM-Arkiv i nuläget (år 2011).

I figur 3 visas hur leveranser av analogt (papper, foto och liknande) och digitalt material i dagsläget (år 2011) ser ut på en övergripande nivå. I framtiden räknar TAM-Arkiv med att få ett systemstöd för långtidslagring av digitalt material. Oberoende om det finns ett systemstöd eller inte i framtiden så kommer leveransproceduren att vara mycket lika den nuvarande modellen.

TAM-Arkiv har ingen möjlighet att ta ansvar för inkommen digital information, om inte TAM-Arkivs instruktioner för digital arkivering följts. TAM-Arkiv kan tyvärr inte garantera bevarande av inkommen digital information om leverans skett utan förberedande kontakt och förberedande arbete utförts. Om en medlemsorganisation exempelvis skickar in en CD-skiva med filer i leverantörsbundna filformat och utan tillhörande dokumentation så kan inte TAM-Arkiv garantera att dessa kommer att vara läsbara i framtiden.

Mycket stora leveranser (exempelvis av digital video eller ljud) har TAM-Arkiv idag ingen teknisk kapacitet att förvalta aktivt. Detta måste istället oftast lagras passivt på band eller optisk disk.

Arbetssteg som görs vid medlemsorganisation:

1. Maila, besök eller ring ansvarig för digital arkivering vid TAM-Arkiv. Se kapitel ”Kontakt”.
2. Bestäm hur framtida åtkomst ska kunna ske.
 - 2.1. Ska informationen vara både läsbar och maskinellt återanvändningsbar, eller bara läsbar för ett mänskligt öga? Vilka format ska vara tillgängliga?⁴
 - 2.2. Vad ska gå att söka utifrån (exempelvis datum, ärendenummer, ärenderubrik...)?
3. Ska informationen endast lagras passivt i arkivdepå (ingår i medlemsavgiften) eller önskas aktiv lagring i system med backup? Aktiv lagring ger en högre informationssäkerhetsnivå, men detta skapar också merkostnader för TAM-Arkiv så i detta fall kan exempelvis en leveransavgift behöva tas ut.
4. Tag en arbetskopia av det som ska levereras. (Detta görs för att inte riskera att förstöra originalinformationen vid bearbetning inför leverans.)
5. Avgränsa det som ska levereras (och gallra bort det som inte behöver levereras).
6. Dokumentera själva informationen och kommande leverans. Dokumentation ska förklara för ”vem-som-helst” vad som ingår och hur leveransen skett.
7. Konvertera till arkivformat.
8. Tillför beskrivande information (metadata) till information som ska levereras om detta behövs. Se till att all metadata till leveransen är framtagen och i rätt format.
9. Paketera informationen som leveranspaket.
10. Leverera på överenskommet sätt.

Arbetssteg som görs vid TAM-Arkiv:

1. Initieras efter kontakt från medlemsförbund: Förbered leveransmottagande.
2. Har TAM-Arkiv teknisk kapacitet att ta emot de kvantiteter som önskas levereras? (Exempelvis video, ljud eller ärendehanteringssystem med stor mängd filer.)
3. Tag emot, packa upp och kontrollera inkommen leverans mot inkommen dokumentation.
4. Begär eventuell komplettering.
5. Bekräfta att leveransen kommit in via email.
6. Lägg i karantän i 30 dagar i isolerad dator. (Detta görs för att minska risken för virusspridning.)
7. Importera till filsystem eller lagra passivt i traditionell arkivdepå.
8. Skicka slutkvittens om att TAM-Arkiv tar över förvaltningsansvaret för informationen.

Arbetssteg som görs vid TAM-Arkiv under framtida informationsförvaltning (gäller endast för aktivt lagrad information):

⁴ Med maskinellt återanvändningsbar menas att information lätt ska kunna återanvändas, utan att behöva matas in manuellt igen, eller OCR-tolkas. Om exempelvis ett kontorsdokument lagras i TIFF-format är det läsbart, men inte maskinellt användningsbart utan OCR-tolkning.

1. Underhåll av information som lagras aktivt. Kontroll-läs, migrera och konvertera vid behov.

Notera att information som endast lagras passivt i arkivdepån underhålls bara om tid och resurser finns tillgängliga vid TAM-Arkiv.

Kompletterande villkor och mer utförliga råd för digital leverans ges i följande avsnitt "Detaljerade villkor för leverans" i denna instruktion.

6.2 Detaljerade villkor för leverans

6.2.1. Bestäm vad som ska levereras

Information som ska gallras ska inte levereras till TAM-Arkiv, om inte annat överenskommit. Riktlinjer för vad som ska gallras och bevaras finns i levererande organisations dokumenthanteringsplan. Om en dokumenthanteringsplan inte finns, eller inte är aktuell, bör en gallringsutredning göras där det beslutas vad som ska gallras respektive bevaras.

Avgränsa den information som ska levereras, oberoende av i hur många system den kan vara lagrad. Fokusera på information och inte system, eftersom det är information som lagras och inte system.

Informationen från detta avsnitt dokumenteras i överenskommelsen mellan levererande part och TAM-Arkiv.

6.2.2. Villkor för framtida åtkomst och söklistor

För att kunna hitta och se levererad information i framtiden måste man bestämma hur man ska kunna återsöka och se den.

Exempel på frågor:

- Hur ska den information som ska levereras kunna återsökas i framtiden? Efter vilka kriterier och sökvillkor? Behöver söklistor skapas?
- Hanterande av eventuellt konfidentiellt material och behörighetskontroll. (Vem får åtkomst? Hur ska kontroller ske?)
- Hur ska informationen presenteras?⁵ Behövs stilmallar eller annan teknisk lösning? Behöver rapporter eller andra former av sammanställningar genereras? Ska leveransen ske i digital form eller på papper?

För att kunna återsöka information bör metadata tillföras för exempelvis informationstyp, process, funktion, fil, ärende eller akt. Filer behöver beskrivas för att de ska lättare kunna konverteras, återsökas och återanvändas. Filmetadata kan exempelvis beskrivas med PREMIS, Dublin Core XML, UK eGov XML ERMS eller något annat XML-format.⁶

Informationen från detta avsnitt dokumenteras i överenskommelsen mellan levererande part och TAM-Arkiv.

⁵ Ursprunglig presentation brukar vara mycket svår att behålla vid långtidslagring.

⁶ TAM-Arkiv hoppas att i framtiden få resurser för att genomföra ett arbete kring vilket metadata som kommer att krävas för att klara långtidslagring, se även tidigare förslag till projektet "TAMAD-Metadataramverk". I väntan på detta så varierar den praktiska lösningen från fall till fall. Även rekommendation för stadier för livscykel för ärenden/dokument skulle behöva göras.

6.2.3. Leveransprocedur

Beskriv hur leveransproceduren ska se ut. Hur ofta ska leveranser ske? På vilket sätt?

En provleverans (innan den skarpa leveransen) brukar vara vettigt att använda för att testa ny systemfunktionalitet eller ett nytt leveranssätt. En provleverans brukar bestå av en mindre informationsmängd, som används för att endast testa export, överföring och import.

Notera att det är först när TAM-Arkiv godkännt den fullständiga skarpa leveransen som TAM-Arkiv tar över förvaltningsansvaret för det levererade materialet. Vänta därför med att förstöra originalmaterialet hos levererade part tills TAM-Arkiv lämnat detta godkännande, för att inte riskera att material försvinner.

Informationen från detta avsnitt dokumenteras i överenskommelsen mellan levererande part och TAM-Arkiv.

6.2.4. Juridiska villkor

Generellt är det öppenhet som gäller vid åtkomst till tidigare levererad digital information. Det vill säga fri tillgång ges till tidigare information. Öppenhet gäller dock inte om ett förbund vid leverans begärt att vissa serier inte ska lämnas ut. Öppenhet gäller inte heller om TAM-Arkiv bedömer att ett utlämnande kan stå i strid med svensk lagstiftning eller om det finns risk att en enskild kan lida skada. Exempelvis så lämnas inte arbetsskadeakter, enskilda löneuppgifter och liknande personrelaterad information ut till forskare.

Information om tillgänglighet till levererat material ska begränsas dokumenteras i överenskommelsen mellan levererande part och TAM-Arkiv. Uppgift om vilken roll (ofta kanslichef eller arkivansvarig) vid levererande medlemsorganisation som kan ge tillstånd till att en forskare får tillgång till konfidentiellt material behöver dokumenteras.

I överenskommelsen dokumenteras även om material ska levereras som 1) gåva (d.v.s. att ägandeskapet för materialet övergår till TAM-Arkiv), eller som 2) deposition (d.v.s. att ägandeskapet för material blir kvar hos levererande part, men det förvaras på TAM-Arkiv).

Ytterligare något som behöver dokumenteras vid leverans är om materialet omfattas av copyright, vem som innehar copyright, tidpunkt då copyright upphör och villkor för copyright.

6.2.5. Finansiering vid leverans

De faktiska extra arbetskostnaderna som TAM-Arkiv kan få vid leverans uppskattas enligt gällande prislista. Exempelvis kan det gälla att en leverans innehåller filer i filformat som är leverantörsberoende och som behöver konverteras till arkivformat.

Finansiering av löpande förvaltning sker enligt gällande prismodell.

Informationen från detta avsnitt dokumenteras i överenskommelsen mellan levererande part och TAM-Arkiv.

6.2.6. Leveransdokumentation

TAM-Arkiv skickar inför varje leverans eller serie av leveranser en XML-fil "PackageMD.xml". I denna fil följer med uppgifter såsom leveransID, arkivbildarID och arkivbildarnamn. Övriga uppgifter såsom exempelvis kontaktuppgifter och paketID kompletteras i filen av levererande medlemsorganisation.

Även innehållet i leveransen behöver dokumenteras för att det ska gå att förstå den i framtiden. Exempelvis behöver koder och använda förkortningar förklaras i klartext. Denna dokumentation ska medfölja leveransen i PDF/A-format eller i annat lämpligt arkivformat.

6.2.7. Leveransmedia

TAM-Arkiv kan ta emot digitala leveranser på USB-flash-minne, extern hårddisk med USB-anslutning, CD-R, DVD-R och DVD+R.

Ytterligare alternativ är att skicka som e-post eller gå via någon filöverföringstjänst som www.dropbox.com, www.opendrive.com, www.sprend.se (www.skickafil.se), www.yousendit.com, www.filecentral.se eller www.storegate.se

Notera att om information bedöms som känslig kan den paketeras och krypteras. Kontrollsummor kan också användas. Lösenord kan sedan ges till TAM-Arkiv muntligen via telefon eller SMS. Vid ankomst till TAM-arkiv upp-packas och avkodas informationen, sätts i karantän och kontrolleras efter skadlig kod under 30 dagar.

Informationen från detta avsnitt dokumenteras i bilaga till överenskommelsen mellan levererande part och TAM-Arkiv.

6.2.8. Leveranspaketstruktur

ZIP/TAR-container eller mapp



Fig 4. Paketstruktur för åtkomst och leveranser.

Digitala leveranser till TAM-Arkiv sker i leveranspaket. Detta avsnitt beskriver hur leveranspaket till TAM-Arkiv ser ut.

- Mapp: "Arkivbildarnummer_arkivbildarförkortning_leveransnummer"
 - Fil: "PackageMD.xml"
 - Filer: Innehållsdokumenterade filer
 - Filer: Eventuella genererade sökindex som behövs för återsökning
 - Mapp: "Data"
 - Filer och mappar i leverans

Fig 5. Filträd i ett leveranspaket

Ett leveranspaket avgränsas i en 1) okomprimerad mapp/katalog, 2) i en ZIP- eller 3) i en TAR-container. Data läggs i en undermapp. Se figur 2 och 3.

- Mapp: "1023_TestForbundet_2981"
 - Fil: "PackageMD.xml"
 - Filer: "Kodförklaring.pdf", "Informationsmodell.pdf"...
 - Fil: "Sökregister_personnr.xml"
 - Mapp: "Data"
 - Mapp: "Pension"
 - Filer: "Pension1.xml", "Pension2.xml", "Pension3.xml"...

Fig 6. Exempel på filträd

Nedanstående metadata behöver anges i leveransfilen "PackageMD.xml":

- PacketID
- LeveransID och leveransdatum
- ArkivbildarID och arkivbildarnamn
- Leveranskontaktuppgifter (vid arkivbildare eller leverantör)
- Ev. titel på leveransinnehållet
- Ev. beskrivning av leveransinnehållet. Ex. historik, kontext, kopplingar till andra system eller leveranser.
- Tidsperiod då informationen skapades/rör (from - tom)
- Ev. konfidentiellt material. Finns uppgifter som inte ska vara allmänt tillgängliga?
- Ev. copyright.
- Ev. anmärkningar
- Ev. vårdlogg (utförd aktivitet, tidpunkt, aktör)

Senaste versionen av exempelfil för leveransdata finns att tillgå i XML-format, se filen "PackageMD.xml" som fås av TAM-Arkiv vid leveransförberedelse.

6.2.9. Filformat och teckentabeller

Tillåtna format för leverans enligt TAM-Arkivs rekommendation "TAM5:2010-Format för långtidslagring". Formatkonvertering sker innan leverans. TAM-Arkiv kan bistå med förslag på tekniska lösningar. Om formatkonvertering inte sker innan leverans kan TAM-Arkiv i en del fall utföra konvertering emot betalning.

Behövs parallell lagring i olika format för att se till att information är både läsbar och återanvändningsbar? Exempelvis kan detta vara aktuellt för CAD, teckningar eller overhead-material.

Informationen från detta avsnitt dokumenteras i bilaga till överenskommelsen mellan levererande part och TAM-Arkiv.

6.2.10. Databaser inklusive register

Databaser finns av många olika typer, exempelvis hierarki, relation, nätverk, objekt, objektrelaterad, träd- och tabellbaserad. För att tolka data i en databas behövs beskrivande data, metadata.

TAM-Arkiv tar emot databaser enligt nedanstående alternativ:

Alt. 1: För MS Access, Oracle och MS SQL-Server: Lagring enligt SIARD-formatet. I detta fall finns det ett enkelt färdigt IT-verktyg som kan användas för att exportera ut data, editera metadata och återskapa en databas.

Alt. 2: Denormaliserad lagring till XML-format, om data inte är tänkt att hanteras i en databas i framtiden. Exempelvis kan detta gälla löneutbetalningar. Ur databasen skapas ett urval av rapporter/vyer och liknande. Lagringen sker om möjligt i standardiserade XML-format.

Alt. 3: Normaliserad lagring i XML-format, om data är tänkt att kunna bearbetas i ”databasform” i framtiden. Exempelvis kan detta gälla forskningsdata. Lagringen sker om möjligt i standardiserade XML-format.

Om man inte kan eller vill använda SIARD-formatet, kan man överväga att använda ADDML version 8.2 eller senare. ADDML har ursprungligen tagits fram vid norska riksarkivet, men förvaltas numera gemensamt av norska och svenska riksarkiven. Om man använder ADDML, så lagras innehåll från databasen som sekventiella ("flata") filer, och metadata om struktur och innehåll lagras i en ADDML-fil. Formatet ADDML kan även hantera binära filer/dokument som länkats till databasen som ska lagras. ADDML är inget verktyg, utan bara ett lagringsformat för metadata till databaser som lagrats som sekventiella filer och tillhörande eventuella objektfiler. Svenska riksarkivet tillhandahåller ett IT-verktyg kallat "RALF", som kan skapa ADDML-filer i ett METS-paket utifrån sekventiella filer. RALF utför även olika former av data-kontroller av de sekventiella filerna man vill leverera. Innan man använder RALF måste man dock först ta ut data ur den databas man vill leverera som sekventiella filer. Både svenska och danska riksarkivet tillhandahåller, via sina respektive hemsidor, verktyg som kan vara användningsbara för att hantera teckenkonvertering och liknande.

Om varken SIARD eller ADDML kan användas så kan man exempelvis utgå från en anpassning av ett tidigare ISO-arbete från år 2000, "A JDBC-oriented XML DTD for SQL result sets". Om inget annat fungerar så får man i värsta fall titta på leverantörsspecifika lösningar, såsom mysqldump-formatet för XML. Dessa former av uttag behöver dock kompletteras med metadata i XML-format.

Vid uttag ur en databas bör man inte ta med de delar som bara hade betydelse under systemets livstid, exempelvis tabeller som innehåller lösenord. Om man använder SIARD för långtidslagring måste man ta en kopia av den ursprungliga databasen, och sedan ta bort eventuella systemtabeller (som inte har någon form av kritisk nyckel-koppling till andra tabeller) ur databaskopian, innan man gör ett uttag till SIARD-formatet.⁷

I många fall förekommer även binära filer, i anslutning till databasen. Dessa filer kan antingen vara 1) inbäddade direkt i databasen som BLOB (Binary Large Object) eller CLOB (Character Large Object) eller 2) vara länkade till databasen. Det senare fallet är vanligare för större filer och i detta fall läggs filerna i en separat filstruktur. Man måste göra en bedömning om man antingen vill ta med de filer som ligger i den separata filstrukturen, eller om man väljer att inte ta med dessa. Om man väljer att ta med filerna måste dessa konverteras till arkivformat. Även länkarna behöver bearbetas om till relativa länkar med aktuella och genomtänkta filnamn. Se TAM-Arkivs rekommendation ”TAM 5:2010 REKOMMENDATION - Format för Långtidslagring”.

⁷ Detta eftersom SIARD-verktyget tar alla tabeller och fält rakt av och exporterar ut dem till SIARD-formatet.

När man lagrar data från en databas i XML-form, kan man lagra data i attribut eller i element.

Alt 1. Lagra i attribut

Alt 1a. Alla specialtecken & < > ' ' omvandlas till < > & " '

Alt 1b. Alla specialtecken & < > ' ' omvandlas till < > " ' &

Märk att inga CDATA-sektioner är tillåtna i attribut..

Alt 2. Lagra i element

Alt 2a. Omvandla alla specialtecken & < > ' och " till < > & " och '

Alt 2b. Lägg allt i CDATA-sektioner.

Använd sekvensen <![CDATA[placera inkapslat data här]]> för att kapsla in data.

Notera att om man misstänker, eller vet, att det kan finnas sekvensen]]> (så kallat CDEnd) i databasen så måste alla eventuella sekvenser av]]> i databasen omvandlas till sekvensen]]]><![CDATA[>⁸

Alt 2c. Detta kräver att alla & < > " omvandlas till < > " ' &

Märk att uttag kan göras på många olika sätt, så därför är det oftast bäst om man kan följa något fastställt standardiserat XML-schema i första hand. Se ”TAM 5:2010 REKOMMENDATION - Format för Långtidslagring” för fler detaljer. Nedan kan ses en principskiss utifrån alternativ 2b och ett enkelt test-schema för relationsdatabaser. Märk att detta är en principskiss och att man kan göra på många olika sätt i praktiken.⁹

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml:stylesheet type="text/xsl" href="TAM_TestDatabase_0_1.xsl"?>
<Set version="0.1" noOfRecordSets="3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="TAM_TestDatabase_0_1.xsd">
  <ExportTime time="22:41:8" date="2004-5-2"/>
  <RecordSet no="1">
    <SQLStatement statement="SELECT * FROM author"/>
    <ColumnsMetaData numberOfColumns="3">
      <ColumnMetaData name="authorID" width="11" dataType="COUNTER" fromSchema=""
        fromTable="authors" nullable="NO" isAutoIncremented="YES" isCaseSensitive="NO"/>
      <ColumnMetaData name="firstName" width="50" dataType="VARCHAR" fromSchema=""
        fromTable="authors" nullable="YES" isAutoIncremented="NO" isCaseSensitive="NO"/>
      <ColumnMetaData name="lastName" width="50" dataType="VARCHAR" fromSchema=""
        fromTable="authors" nullable="YES" isAutoIncremented="NO" isCaseSensitive="NO"/>
    </ColumnsMetaData>
    <Row rowNo="1">
      <Column name="authorID"><![CDATA[1]]></Column>
      <Column name="firstName"><![CDATA[Harvey]]></Column>
      <Column name="lastName"><![CDATA[Deitel]]></Column>
    </Row>
    <Row rowNo="2">
```

8 För den tekniskt intresserade så omvandlas sekvensen]]]><![CDATA[> av XML-parsern till just]]>, därav ersättningen. Testa gärna själv. Ledtråd:]] + CDEnd + CDStart + >.

9 Självkärligt kan attributet ”name” för kolumnerna ersättas med ett sifferindex, ”index” om man föredrar detta.

```

    <Column name="authorID"><![CDATA[2]]></Column>
    <Column name="firstName"><![CDATA[Tem]]></Column>
    <Column name="lastName"><![CDATA[Nieto]]></Column>
  </Row>
</RecordSet>
<RecordSet no="2">
  <SQLStatement statement="SELECT * FROM review"/>
  <ColumnsMetaData numberOfColumns="9">
    ...koden förkortad här...
  </RecordSet>
</Set>

```

För att leverera en databas till TAM-Arkiv krävs följande arbetssteg förutom de mer generella (Notera att om man använder exempelvis SIARD-verktyget så behövs inte alla arbetssteg nedan.):

Steg a: Avgränsa datamängden som ska långtidslagras. Exempelvis kan detta gälla om data är spritt över flera databaser av prestandaskäl.

Steg b: Bestäm om data ska kunna bearbetas i framtiden eller om det är tänkt att i första hand endast kunna läsas i fixerad form. Bestäm hur data ska kunna vara tillgängligt i framtiden. Hur snabbt och på vilket sätt?

Steg c: Bestäm om eventuella filer kopplade till databasen ska långtidslagras och i så fall på vilket sätt.

Steg d: Bestäm vilken informationen i tabeller, kolumner, objekt, attribut som ska långtidslagras. Bestäm hur eventuella sammanställningar/vyer ska se ut.

Steg e: Bestäm vilken teckentabell som ska användas för arkivformatet.

Steg f: Bestäm i detalj vilket XML-Schema som ska följas och om data ska lagras i attribut eller i element.

Steg g: Skapa en kopia av databasen och eventuella separata filer. Använd aldrig "originaldatabasen" för arkivuttag. Detta görs av informationssäkerhetsskäl, om något inte skulle fungera som det var tänkt, vilket händer...

Steg h: Skapa eventuella vyer/rapporter som är sammanställningar från flera tabeller/delmängder. Detta kan gälla både denormaliserad och normaliserad lagring.

Steg i: Konvertera eventuella filer till arkivformat och arbeta om absoluta länkar till relativa länkar med genomtänkta filnamn. Se till att filstrukturen är genomtänkt och klarar långtidslagring. Se även "TAM 5:2010 REKOMMENDATION - Format för Långtidslagring" för fler detaljer.

Steg j: Export enligt något av ovanstående alternativ.

Steg k: Granska och validera exporterat data och komplettera med allt metadata som behövs för att förstå och hantera data i framtiden.

Steg l: Testa återsökning och beställning av exporterat data. Granska resultatet.

6.2.11. Kontorsdokument, fotografier, tidskrifter, CAD-filer och liknande

Fristående kontorsdokument, som placerats i ett filträd (exempelvis på en gemensam eller personlig katalog) bör ha självförklarande namn med versionsnummer, gärna enligt rekommendation "TAM3:2008-Versionshantering". Filer som hör ihop grupperas i en

gemensam katalog, exempelvis efter ärendenummer, projektnamn, dokumenthanteringsplanskod eller liknande. Namn på filer och kataloger bör inte innehålla specialtecken, se även "TAM5:2010-Format för långtidslagring". Exakt lösning varierar något från fall till fall.

För att leverera fristående filer till TAM-Arkiv krävs följande arbetssteg förutom de mer generella:

Steg a: Bestäm vilka filer som ska levereras.

Steg b: Bestäm om data ska kunna bearbetas eller om data bara ska kunna läsas. Om man vill kunna bearbeta filer i framtiden, behöver man spara originalfilen parallellt med en konverterad arkivkopia. Se även "TAM5:2010-Format för långtidslagring" för mer detaljer.

Steg c: Bestäm vilken teckentabell som ska användas för arkivformatet.

Steg d: Skapa en kopia av filträd och fristående filer.

Steg e: Gallra bort eventuella kopior som inte ska bevaras.

Steg f: Konvertera filer till arkivformat och arbeta om länkar till relativa länkar med genomtänkta filnamn. Se till att filstrukturen är genomtänkt och klarar långtidslagring. Se även "TAM 5:2010 REKOMMENDATION - Format för Långtidslagring" för fler detaljer. Spara parallellt i originalformat om filerna ska kunna behandlas i framtiden i originalformat.

Steg g: Granska exporterat data och komplettera med allt metadata som behövs för att förstå och hantera data i framtiden.

Steg h: Testa återsökning och beställning av exporterat data. Granska resultatet.

6.2.12. E-post

E-post sparas som 1) rena textfiler, 2) i XML-format, 3) i XHTML/HTML-format eller i PDF/A-format. E-post kan även sparas som poster i en databas och långtidslagras då på samma sätt som en databas, se ovan.

I de flesta fall kan email som långtidslagras bara sökas ut efter rubrik, avsändare, mottagare och avsändartidpunkt.¹⁰

För att leverera fristående email till TAM-Arkiv krävs följande arbetssteg förutom de mer generella:

Steg a: Bestäm vilka email som ska levereras. De flesta email brukar ha lågt bevarandevärde. Email väljs primärt ut utifrån avsändare, mottagare och sändtidpunkt.

Steg b: Bestäm om bifogade filer ska långtidslagras. Bestäm även om bifogade filer ska kunna bearbetas i framtiden eller om dessa bara ska kunna läsas. Om man vill kunna maskinellt bearbeta bifogade filer i framtiden, behöver man ofta spara originalfilen i ursprungsformat parallellt med en konverterad arkivkopia för långtidslagring. Se även "TAM5:2010-Format för långtidslagring" för mer detaljer.

Steg c: Bestäm vilken teckentabell som ska användas för arkivformat och om emailen ska lagras som ren text, XML, XHTML, HTML, PDF/A-1 eller på annat sätt.

¹⁰ Dessutom finns det mer teknisk information som berättar vägen som emaillet har tagit, vilken epost-klient som använts, protokoll och liknande, men denna information brukar inte efterfrågas av den vanlige användaren.

Steg d: Exportera ut email och eventuella bifogade filer till en filstruktur (eller till en databas för långtidslagring).

Steg e: Gallra bort eventuella kopior som inte ska bevaras.

Steg f: Konvertera filer till arkivformat och arbeta om länkar till relativa länkar med genomtänkta filnamn. Se till att filstrukturen är genomtänkt och klarar långtidslagring. Se även "TAM 5:2010 REKOMMENDATION - Format för Långtidslagring" för fler detaljer. Spara parallellt i originalformat om filerna ska kunna behandlas i framtiden i originalformat.

Steg g: Granska exporterat data och komplettera med allt metadata som behövs för att förstå och hantera data i framtiden.

Steg h: Testa återsökning och beställning av exporterat data. Granska resultatet.

6.2.13. Diarium och register

Ett diarium eller ett register lagras som en databas, se avsnitt ovan.

6.2.14. Dokumenthanterings- och ärendehanteringssystem

Dokumenthanteringssystem (DMS, Document Management System), ärendehanteringssystem, innehållshanteringssystem (CMS, Content Management System), portaler och liknande kan erbjuda många olika vyer mot samma bakomliggande data. I ett dokumenthanterings- eller ett ärendehanteringssystem kan informationen vara lagrad i flera olika databaser och filsystem. Ofta är dessa system dessutom integrerade med intra- och extranät, se även avsnittet om webbsiter nedan.

I princip så långtidslagras data på samma sätt som för databaser och fristående filer, men några tillägg:

Tillägg 1. Gör först en avgränsning av vilka informationsmängder som ska långtidslagras, oberoende av hur de råkar vara tekniskt lagrade i olika system/databaser.

Tillägg 2. Bestäm om data ska tas ut som denormaliserade vyer¹¹ eller om det är mer intressant att bevara bakomliggande data bakom vyerna. Konstruera vid behov nya vyer mot bakomliggande data. Detta kan vara riktigt komplext om informationsarkitekturen är ogenomtänkt eller dåligt dokumenterad.

Tillägg 3. Bestäm om data ska kunna bearbetas i framtiden eller om det räcker med att bara kunna återsöka och läsa det. Om data ska kunna bearbetas på nytt i framtiden, behöver rådata lagras parallellt med eventuella skapade vyer av denormaliserat data. Även ursprungliga format kan behöva sparas parallellt med arkivformat, om långtidslagrat data ska kunna bearbetas.

Tillägg 4. Bestäm hur återsökningar ska kunna ske. Nya söklistor, sökdata och index kan behöva skapas. Hur ska informationen kunna återsökas?

För övrig information om långtidslagring av dessa former av system se avsnitten om databaser, kontorsdokument, email och webbsiter.

¹¹ Exempel på en denormaliserad vy kan vara en enkel lönelista, som bakom kulisserna kan vara sammanfogad utifrån ett flertal olika källor (ex. relationstabeller).

6.2.15. Webbiter, bloggter, sociala media och källkod

Långtidslagring av webbiter ("hemsidor") påminner mycket om långtidslagring av dokument- och ärendehanteringssystem. Att långtidslagra webbiter är oftast relativt komplext och det finns än så länge inga enkla lösningar för dynamiska siter som fungerar i alla situationer. Grundproblemet är att de flesta webbiter konstrueras utan tanke på att dessa skulle kunna behöva långtidslagras i framtiden.

Det finns i princip ett antal olika möjliga principlösningar för att långtidslagra en dynamisk webbplats, exempelvis:¹²

Alt. 1. Omvandla ett snapshot (en "ögonblicksbild") av hela eller delar av webbsiten vid ett specifikt tillfälle till statiska HTML/XHTML-sidor med en statisk filstruktur, eller till ett eller flera PDF/A-dokument. (Om man har tur och den ursprungliga webbsiten mot förmodan redan är statisk behöver inte detta göras förstås...:-)

Alt. 2. Omvandla webbsiten till artiklar i PDF/A, XML, HTML eller XHTML-format med metadata i XML (eller i en databas). Exempelvis webbsiten Vårdguiden använder detta alternativ.

Alt. 3. Lagra ett urval ur bakomliggande databaserna och filsystem separat. Bortse från den ursprungliga presentationen, det vill säga lagra endast ursprungligt bakomliggande rådata.

TAM-Arkiv kan ta emot webbiter enligt någon av de tre ovanstående principlösningarna, eller kombination av dessa. Vilken principlösning man än väljer så får man nästan alltid en informationsförlust, då det inte går att bevara ursprunglig funktionalitet och presentation.

För mer information om lagring av webbiter, så kan man exempelvis studera The Internet Archive (<http://www.archive.org/>). I Sverige har bland annat Kungliga biblioteket, LDB-centrum¹³ och Riksarkivet prövat olika metoder för att lagra webbplatser.

Vid långtidslagring av webbiter används samma arbetssteg som vid långtidslagring av databaser och kontorsdokument, se ovan. Exempelvis innebär det att filer måste konverteras till arkivformat, absoluta länkar måste göras om till relativa länkar, exekverbara filer måste tas bort och så vidare. Se även "TAM 5:2010 REKOMMENDATION - Format för Långtidslagring" för fler detaljer runt webbformat. Länkarna mellan sidor måste också kontrolleras och ibland repareras.

Något som man också bör tänka på är att exempelvis klientscript helst ska tas bort och ersättas av statiska HTML-, XHTML- eller PDF/A-sidor. Detta eftersom funktionalitet för

12 Det finns även andra alternativ som utgår från krav på att informationen på en webbplats måste kunna återskapas i ursprungligt vid ett specifikt tillfälle av juridiska skäl, men de är ofta mindre lämpliga för långtidslagring, ett exempel på detta är "fullständig http-loggning".

13 Se exempelvis LDB-centrums CAST (Collaborative Archiving Services Testbed) och tidigare rapporter om CODA-WEBB.

program-/scriptkod inte går att bevara på lång sikt.¹⁴ Alla former av exekverbara filer måste också tas bort.

Notera att TAM-Arkiv varken garanterar eller vårdar kopior på inskickade webbsiter som inte omvandlats till arkivformat och arkivstruktur. För att TAM-Arkiv ska kunna vårda och tillhandahålla en webbplats i framtiden krävs att den omvandlas till arkivformat och arkivstruktur, och att den inte innehåller exekverbara filer. Om en leverans innehåller exekverbara filer så har inte TAM-Arkiv möjlighet att vårda eller tillgängliggöra denna leverans. Med andra ord den kastas eller ställs den på en hylla i den traditionella arkivdepån. Detta av säkerhetsskäl och därför att det inte går att garantera rätt fungerande kodfunktionalitet på lång sikt.

Däremot kan TAM-Arkiv ta emot ”källkod” (d.v.s. okompilerad, ej exekverbar programkod), om någon skulle vilja leverera detta till TAM-Arkiv som en del i systemdokumentation eller motsvarande.

¹⁴ Det är relativt vanligt med krånglande script redan när en webbsida ligger på den ursprungliga webbplatsen. Hur stor är sannolikheten att ett script fungerar om femtio år?

6.3 Revisionshistorik

<i>Revision/Version</i>	<i>Datum</i>	<i>Författare</i>	<i>Kommentar</i>
'1.0'	'2011-09-01'	Magnus Wåhlberg	Första revisionen släpptes. Totalt 25 versioner skapades innan första revisionen fastställdes.

6.4 Länkar

Svenska Riksarkivets dokumentation över och anpassningar av XML-vokabulärer såsom ADDML, BTLA(SIEXML), EAC, EAD, METS, MIX, PREMIS:

<http://xml.ra.se/>

Svenska Riksarkivets verktyg för leveranskontroll och arkivuttag vid leveranser av databaser (RALF m.m.):

<http://www.riksarkivet.se/default.aspx?id=4550&refid=1691#RALF>

LDB-Centrum, för utredningar runt långtidslagring, verktyg, format et cetera:

<http://www.ltu.se/centres/Centrum-for-langsiktigt-digitalt-bevarande-LDB>

Danska Statens Arkiver ("Danska riksarkivet"):

<http://www.sa.dk>

Danska Digitalbevaring.dk:

<http://digitalbevaring.dk/>

Norska riksarkivets sida om ADDML:

<http://www.arkivverket.no/arkivverket/Arkivbevaring/Elektronisk-arkivmateriale/Standarder/ADDML>

Schweiziska Riksarkivets SIARD-sida:

<http://www.bar.admin.ch/dienstleistung/00823/00825/index.html?lang=de>

Paradigm:

<http://www.paradigm.ac.uk/workbook/>

Digital preservation vid Library of Congress:

<http://www.digitalpreservation.gov/>

Exempel på EU-projekt som arbetar/arbetat med frågor runt långtidslagring:

CASPAR (<http://www.casparpreserves.eu/>)

Planets (<http://www.planets-project.eu/>)

Open Planets Foundation (<http://www.openplanetsfoundation.org/>)

Digital Preservation Europe (<http://www.digitalpreservationeurope.eu/>)

6.5 TAM-Arkivs rekommendationer och rutiner

Följande rekommendationer och rutiner kan laddas ner från TAM-Arkivs webbplats, <http://www.tam-arkiv.se>:

TAM 1:2008 REKOMMENDATION - Vid förändring av ärende- och dokumenthantering

TAM 2:2008 REKOMMENDATION - Informationsförvaltning

TAM 3:2008 REKOMMENDATION – Versionshantering

TAM 4:2010 REKOMMENDATION - Arkivfunktionalitetskrav för IT-system

TAM 5:2010 REKOMMENDATION - Format för Långtidslagring

TAM 6:2010 REKOMMENDATION – Struktur för dokumenthanteringsplan

TAM 7:2011 INSTRUKTION - Komplet teknisk instruktion för digital åtkomst och digitala leveranser till TAM-Arkiv