



Titel: Rekommendation – Format för långtidslagring

Datum: 2010-03-11

Version 1.0

TAM 5:2010 REKOMMENDATION

Format för långtidslagring

Version 1.0

© Copyright TAM-Arkiv 2010.

Detta verk får kopieras och distribueras fritt i sin helhet så länge som källan anges.

TAM-Arkiv
Grindstuvägen 48
167 33 Bromma

Innehållsförteckning

1. FÖRORD	6
2. METOD, URVALSPRINCIPER, DISPOSITION OCH KÄLLOR	8
3. ORGANISATIONER SOM ARBETAR MED DIGITAL ARKIVERING	10
4. DATAFORMAT	11
4.1 ALLMÄNT	11
4.2 REKOMMENDATIONER	11
4.2.1 DATUM OCH KLOCKSLAG	11
4.2.2 FYSISKA PERSONERS NAMN	12
4.2.3 PERSONNUMMER	12
4.2.4 SAMORDNINGSNUMMER	12
4.2.5 ORGANISATIONSNUMMER	13
4.2.6 POSTADRESS	13
4.2.7 POSTNUMMER	13
4.2.8 GEOGRAFISKA KOORDINATER	13
4.2.9 SPRÄKKODER	14
4.2.10 LANDSKODER OCH SVENSKA LÄNSKODER	14
4.2.11 VALUTAKODER	15
4.2.12 BANKKONTORSNUMMER	15
4.2.13 BANKKONTONUMMER	15
4.2.14 ARKIVSERIEID FÖR PUNKTNOTATION	15
4.2.15 FIL- OCH KATALOGNAMN	15
5. TECKENTABELLER	16
5.1 ALLMÄNT	16
5.2 REKOMMENDATIONER	16
5.2.1 WESTERN LATIN-1 (ISO-8859-1)	16
5.2.2 UTF-8 OCH UTF-16 (UNICODE TRANSFORMATION FORMATS)	17
6. FILFORMAT	18

6.1 ALLMÄNT	18
6.2 REKOMMENDATIONER	20
6.2.1 W3C XML OCH ISO 8879:1986 SGML	20
6.2.2 PDF/A-1 (ISO 19005)	21
6.2.3 ODF (ISO 26300)	22
6.2.4 LEVERANTÖRSBEROENDE KONTORSFORMAT	22
6.2.5 EMAIL-FORMAT	23
6.2.6 WEBBFORMAT (XML, XHTML, HTML, CSS, XSLT, XSL-FO...)	23
6.2.7 JPEG (ISO 10918)	24
6.2.8 JPEG 2000 (ISO 15444)	25
6.2.9 PNG (ISO 15948:2004)	25
6.2.10 GIF 89A (GRAPHICS INTERCHANGE FORMAT)	25
6.2.11 TIFF	26
6.2.12 SVG (W3C SVG)	26
6.2.13 WEBCGM OCH CGM (ISO/IEC 8632:1999)	26
6.2.14 U3D (ECMA 363)	26
6.2.15 PDF/E (ISO 24517-1:2008)	27
6.2.16 X3D (ISO 19775 M.FL.)	27
6.2.17 GML (ISO 19136:2007)	27
6.2.18 WAVE LPCM	28
6.2.19 MPEG-1 (ISO 11172:1993, ITU-T STANDARD H.261) INKL. MP3	28
6.2.20 MPEG-2 (ISO 13818:1995)	28
6.2.21 MPEG-4 (ISO 14496), DIVX OCH XVID	29
6.2.22 DV OCH DV-DIF	30
6.2.23 VÄRDINFORMATION (DICOM, HL7, HISA...)	30
6.2.24 GENERELLA FORMAT FÖR LAGRING AV TEKNISKT DATA (CDF, HDF...)	31
6.2.25 GENERELLA FORMAT FÖR LAGRING AV DATA UR DATABASER (ADDML, SIARD...)	31
6.2.26 FORMAT FÖR FINANS OCH REDOVISNING (SVEFAKTURA, XBRL...)	32
6.2.27 ARKIVSTANDARDER (EAD, EAC-CPF, PREMIS, METS, MIX, TEXTMD...)	33
6.2.28 GENERELLA METADATASTANDARDER (DUBLIN CORE, RDF, PRISM...)	34
6.2.29 JURIDISK INFORMATION (LEGALXML...)	34
6.2.30 SYSTEMUTVECKLING	35
6.2.31 ARTIKLAR OCH BÖCKER	35
6.2.32 PERSONALRELATERAD INFORMATION	36
6.2.33 DIGITALA SIGNATURER	36
6.2.34 "WRAPPERS"-FORMAT (ASF, ISO_BMFF, RIFF, MXF, WARC, ZIP, TAR...)	37
7. VANLIGA FRÅGOR OCH SVAR (FAQ)	39
7.1 ALLMÄNT	39
7.2 FRÅGOR RUNT FORMAT FÖR LÅNGTIDSLAGRING	41
8. REVISIONSHISTORIK	43

9. BILAGOR	44
9.1 BILAGA 1, ETT URVAL AV KÄLLOR PÅ WORLD WIDE WEB	44
9.2 BILAGA 2, ETT KORTARE URVAL AV RAPPORTER, UTREDNINGAR OCH FÖRESKRIFTER	46
9.3 BILAGA 3, TAM-ARKIVS REKOMMENDATIONER OCH RUTINER	47

1. Förord

En sammanställning av format som lämpar sig för digital arkivering kan aldrig bli komplett. Nya format tillkommer hela tiden och gamla slutar att användas. Livslängden för digitala format är svindlande kort jämfört med livslängden för säg hettitiska lertavlor, medeltida pergamentskrifter eller vikingatida runstenar. Den informationsteknologiska virvelstorm som omger oss numera skapar ständigt ny teknologi inklusive nya lagringssätt. Ur ett historiskt perspektiv är detta ett nytt fenomen som omformar själva grunderna för hur mänskligheten bygger upp sina samhällen och hur man kommunicerar information till framtida generationer eller arter. Det är sannolikt att huvuddelen av informationen som skapas idag aldrig kommer att vara tillgänglig i framtiden, endast rännilar kommer att ha sipprat vidare. Detta är en följd av att information inom organisationer oftast inte förvaltas på ett genomtänkt sätt. Termen ”the digital darkness” (”det digitala mörkret”) brukas användas för att beskriva hur dagens information är på väg in i ett informationsteknologiskt svart hål, där endast informationsrester blir kvar vid händelsehorisonten. Å andra sidan kan man konstatera att den informationsmängd som finns kvar sedan vikingatiden är försvinnande liten, eftersom den till stora delar var verbal.

För att digitalt lagrad information ska kunna vara tillgänglig i framtiden och inte dras ner i det ”svarta hålet” måste den förvaltas löpande. Detta sker genom återkommande och systematisk migrering av information mellan olika tekniska miljöer, inkluderande konvertering mellan olika format. Huvudalternativet till migrering av data brukar vara emulering av ursprunglig miljö, något som teoretiskt sätt är fullt möjligt att göra. I praktiken brukar det dock vara svårt och dyrt att kunna garantera en exakt emulering av en äldre teknisk plattform och programvara. Om tekniska system däremot vore självreparerande och självbeskrivande, så skulle läget vara annorlunda. Exempelvis skulle lagring i DNA/RNA-strukturer, eller med självreparerande nanoteknologiska strukturer, kunna vara något som i framtiden underlättar långtidslagring. Problemet med självbeskrivning skulle man dock fortfarande ha kvar, vilket går tillbaka till ett djupare matematiskt-filosofiskt problem.

Ett annat problem, med kortare tidshorisont, och som rör standarder är att de ofta tolkas ganska olika mellan olika leverantörer. Två leverantörer kan säga att de stödjer samma format men ändå kan deras tolkningar av formatet vara sinsemellan inkompatibla. Det vill säga den ena leverantörens tolkning av ett format går inte att hantera med den andra leverantörens produkt. För att komma ifrån detta problem kan man ta fram ett valideringsverktyg utifrån en gemensamt överenskommen tolkning. Vid denna form av ”conformance testning” kan man direkt se i vilken grad en produkt uppfyller den överenskomna tolkningen av en standard. En annan form av lösning, för att komma fram till hur en standard ska tolkas, är att gemensamt ta fram en öppen referensimplementation. Det vill säga att man bestämmer hur en gemensam teknisk lösning ska se ut, en praktisk tolkning av en viss standard. Referensimplementationen kan sedan användas som ”mall” för hur en standard ska tillämpas. En praktisk och något

cynisk sammanfattning är att man aldrig ska bli förvånad när ett format inte fungerar mellan olika leverantörers produkter. Detta kan gälla såväl dataformat, teckentabeller som filformat. Ett sätt att minska risken för olika tolkningar och missförstånd är att kräva öppenhet från leverantören, dels i dokumentationen av själva formatet, dels den produkt som säger sig stödja formatet. Det vill säga att man använder öppna standarder och dessutom visar hur man tolkar dessa. Att använda öppna standarder ökar sannolikheten att man kan bevara digitalt lagrad information.

Denna rekommendation, TAM 5:2010, är en genomgång av de filformat, teckentabeller och andra dataformat som är lämpade för långsiktig informationshantering. Rekommendationen riktar sig primärt till TAM-Arkivs medlemsorganisationer med syftena att dels hjälpa dessa med att höja kvalitén på sin interna informationshantering och dels beskriva vilka format som är lämpliga för digitala leveranser till TAM-Arkiv. Sekundärt torde den även kunna vara till nytta för andra organisationer som bakgrundsmaterial.

Ett stort tack går till Stig-Lennart Molneryd vid Kungliga biblioteket, Göran Lindqvist och Lena Lindbäck vid LDB-Centrum, Karin Bredenberg och Magnus Geber vid Riksarkivet, Björn Lundell vid Högskolan i Skövde och Bo Johansson vid Sollentuna kommun, vilka alla på olika sätt har kommenterat, visat på brister och gett förslag på förbättringar under olika arbetsstadier. Eventuella kvarvarande sakfel står undertecknad för.

TAM 5:2010 är en av de rekommendationer och rutiner som tagits fram vid TAM-Arkiv och som har till syfte att hjälpa medlemsorganisationerna med deras informationshantering, se även bilaga 3, ”TAM-Arkivs rekommendationer och rutiner”.

Magnus Wählberg
TAM-Arkiv
Alvik i mars år 2010

2. Metod, urvalsprinciper, disposition och källor

Arbetsättet vid framtagande av denna rekommendation har grovt sett bestått av följande moment: bestämmande av urvalskriterier, faktainsamling och urval av format, avstämning med olika experter, sammanställning och presentation. Vissa informella tester har också gjorts av en hel del format såsom UTF-8, CSS, DC-XML, EAC(beta), EAC-CPF, EAD, GIF, JPEG, NewsML, ODF, PDF/A-1, PNG, PRISM, RDF, SVG, TIFF, XHTML, XML, XSLT med flera. Dessutom ligger mångårig personlig praktisk erfarenhet bakom bedömningarna av olika formats värde för långtidslagring.

Skriftliga och muntliga reflektioner över olika format har inhämtats från följande organisationer: Kungliga Biblioteket, LDB-Centrum, Riksarkivet, Högskolan i Skövde och Sollentuna Kommun. Märk att i undantagsfall är det möjligt att bedömningarna i denna rekommendation kan avvika från nuvarande krav som Riksarkivet, Kungliga Biblioteket och Sollentuna kommunarkiv ställer. Detta beror delvis på att TAM-Arkivs uppdragsgivare inte är stat eller kommun, utan finns inom den privata organisationsfären. Denna rekommendation ska därför snarare, i dessa fall, ses som ett komplement till de föreskrifter och regleringar som görs av olika myndigheter verksamma inom området. Detta dokument är primärt ämnat som en vägledning för medlemsorganisationer knutna till TAM-Arkiv, exempelvis inför leverans eller för att förbättra den interna informationskvalitén inom medlemsorganisationen. Sekundärt torde rekommendationen även kunna vara till hjälp för många andra organisationer.

Urvalen har gjorts i två faser: nämligen efter användningsområde och utifrån hur lämpligt formatet är för långtidslagring. De användningsområden som bedömts kunna vara förekommande vid någon medlemsorganisation till TAM-Arkiv i nuläget eller i framtiden har tagits med.

Följande användningsområden har tagits med:

- kontorsformat inklusive format för scanning,
- webb,
- CAD (Computer Aided Design),
- GIS (Geographic Information System),
- vårdjournaler,
- ekonomisk redovisning,
- personalområdet,
- juridik,
- multimedia,
- metadata,
- arkiv,
- systemdokumentation,
- databaser
- och generella dataformat.

Format som exempelvis specifikt rör finansbranschen, försäkringbranschen, industri, CAE, CAM och akademisk forskning har i dagsläget inte tagits med eftersom dessa inte bedömts ha tillräckligt stor betydelse för TAM-Arkivs medlemsorganisationer.

Det finns många olika organisationer som har studerat vilka format som kan fungera vid digital arkivering. Exempelvis Kungliga Biblioteket, Riksarkivet, LDB-Centrum, National Archives i Australien och Library of Congress i USA. Exempelvis LDB-Centrum och Library of Congress har båda två tagit fram urvalskriterier för att bedöma om ett format passar för långtidslagring. Dessa organisationers urvalskriterier verkar välgenomtänkta och överensstämmer begreppsmässigt i hög grad med varandra.

Urvalskriterierna för vilka format som presenteras i detta dokument är en förenkling utifrån de urvalskriterier som tagits fram av Library of Congress och LDB-centrum. Vid urval av de format som presenteras som lämpliga för långtidslagring har en avvägning gjorts mellan:

- komplexitet och intern struktur,
- stabilitet över tiden,
- eventuellt leverantörs-/patentberoende
- och en bedömning av hur spritt formatet är.

Enkla format som förändras sällan eller aldrig, och med en genomtänkt inre struktur, utan leveransberoende och med stor spridning brukar ha större möjligheter till att fungera för långtidslagring än dess motsats. Däremot har ingen speciell hänsyn tagits till om ett format råkar vara ett ”de jure”-format. Exempelvis har flera av ISO-formaten både patent- och leverantörsberoenden.

I kapitel 3 beskrivs kort viktigare organisationer och projekt som arbetar med digital arkivering. Kapitlet därefter beskriver vilka format som bör användas i fält, filnamn, katalognamn och dylikt för att underlätta kommunikation mellan system och för långtidslagring. I kapitel 5 beskrivs de teckentabeller som är mest lämpade för långtidslagring. Därefter beskrivs i kapitel 6 de filformat som fungerar väl för långtidslagring. I detta kapitel beskrivs även kort hanteringen av digitala signaturer, vilket är ett mycket komplext område, och så kallade ”wrappers” som används för paketering och i vissa fall även komprimering. I kapitel 7 finns vanliga frågor och svar runt format och långtidslagring, inklusive enklare förklaringar av vissa använda termer och begrepp. Revisionshistoriken för detta dokument kan ses i kapitel 8.

I bilaga 1 och 2 presenteras ett urval av källorna till den information som presenteras i denna skrift. Redovisningen av källor är dock inte på långa vägar komplett, utan ger endast en fingervisning om var man kan hitta mer information om viktigare format. Notera att ISO tar betalt för kopior av de flesta standarderna som denna organisation förvaltar. Detta i skarp kontrast mot det synsätt som exempelvis W3C och IETF har. Bilaga 3 är en förteckning över de rekommendationer och rutiner som TAM-Arkiv än så länge gett ut i TAM-serien.

3. Organisationer som arbetar med digital arkivering

Det finns ett stort antal nationella och internationella organisationer som arbetar med frågor runt digital arkivering. I Sverige har Riksarkivet en mångårig erfarenhet av digital arkivering, där man har arbetat med format såsom ISO-8859-1, EAD, EAC, SIE-XML, ADDML, TIFF, ERMS, PREMIS och METS. Sedan 1970-talet har man tagit emot leveranser av flata filer från olika former av databaser. Notera att Riksarkivet bestämmer hur arkivhanteringen ska ske inom statlig sektor. Riksarkivet har dessutom en närmast normativ roll för uttolkningen av exempelvis EAC, EAD och ADDML i Sverige, se <http://xml.ra.se/>.

Kungliga Biblioteket har länge arbetat med digital arkivering av exempelvis webbsidor. Audiovisuella medier (hette tidigare Statens ljud- och bildarkiv) som numera ingår i Kungliga Bibliotek, har omfattande kunskaper inom digital arkivering av ljud och film. Kungliga Biblioteket har även studerat digital arkivering av artiklar och böcker.

LDB-Centrum arbetar med forskning och utveckling inom digital informationshantering. Bland annat har man arbetat med digital arkivering av webbplatser och information från ärendehanteringssystem. LDB-centrum har i samarbete med andra organisationer tagit fram ett antal rapporter som är läsbara och öppet tillgängliga, se Bilaga 1 och 2.

På landstingskommunalnivå finns det ett antal landstingsarkiv som arbetar med frågor runt digital arkivering. Även på kommunal nivå har allt fler kommun- och stadsarkiv börjat arbeta med frågor runt digital arkivering. Hur arkivhanteringen styrs kan variera något mellan olika kommuner och landsting, men oftast brukar respektive kommunarkiv/landstingsarkiv ha fått mandat att styra arkivhanteringen inom respektive kommun/landsting.

Dessutom har allt fler statliga myndigheter och privata organisationer satt upp olika former av IT-system som används för långtidslagring av information. Vid konstruktionen av dessa system har man ofta inspirerats av NASA:s modell OAIS (Open Archival Information System), även om man i de flesta fall inte följt den fullt ut.

Internationellt finns det en mycket stor mängd organisationer som arbetar med digital arkivering. De flesta riksarkiv och nationella bibliotek i den industrialiserade världen är numera mer eller mindre aktiva i frågan. Exempelvis riksarkiven i USA, England, Australien, Nederländerna, Portugal och Schweiz. Library of Congress och NASA brukar också ofta lyftas fram som två av de ledande aktörerna inom området. Om man vill ha en överblick runt vad som händer inom digital arkivering och hur man ska gå till väga kan "Paradigm – Workbook on Personal Digital Archives" framtagen i ett samarbete mellan universiteten i Oxford och Manchester vara en bra utgångspunkt. Se även bilaga 1.

Inom EU finns för tillfället minst fyra stora projekt som arbetar med forskning och utveckling runt digital arkivering, nämligen Caspar, Digital Preservation Europe, Planets och Protage. De olika EU-projekten har lite olika inriktning, metoder, resurser och inblandade organisationer, se bilaga 1. Det finns omfattande material att hämta från deras respektive webbplats, inklusive vissa verktyg.

4. Dataformat

4.1 Allmänt

För att få en fungerande långtidslagring måste man använda enhetliga och välgenomtänkta dataformat i fält, variabler, poster och filnamn. I detta kapitel beskrivs vanligt förekommande dataformat som man bör använda oberoende av tidsperspektiv.

Ingen beskrivning av dataformat kan dock vara komplett, med tanke på att det finns en stor flora av olika rekommendationer och standarder för detta område. Bland de standarder som exempelvis inte beskrivs på grund av omfattningen är ISO 11179 ”Specification and Standardization of Data Elements, Parts 1-6”, som kan användas för att beskriva dataelement.

4.2 Rekommendationer

4.2.1 Datum och klockslag

W3C:s begränsning utifrån ISO 8601:1988 (E), “Data elements and interchange formats - Information interchange - Representation of dates and times”, beskriver hur datum och tid ska lagras. År 2005 tog W3C även fram ytterligare rekommendationer för hanteringen av tidszoner, se ”Working with Time Zones - W3C Working Group Note 13 October 2005”.

Vanliga exempel på lagring av tid:

År

YYYY (ex. 1997)

År och månad

YYYY-MM (ex. 1997-07)

Komplett datum

YYYY-MM-DD (ex. 1997-07-16)

Komplett datum med timmar, minuter och sekunder

YYYY-MM-DDThh:mm:ss (ex. 1997-07-16T19:20:30)

Komplett datum med timmar, minuter, sekunder och tidszon

YYYY-MM-DDThh:mm:ssTZD (ex. 1997-07-16T19:20:30+01:00)

där

YYYY = år, MM = månad (01 = januari et.c.), DD = dag (01 till och med 31), hh = timme (00 till och med 23), mm = minut (00 till och med 59), ss = sekund (00 till och med 59), T = avskiljare mellan år och datum, TZD = tidszonbeskrivning (Z eller +hh:mm eller -hh:mm).

Vid presentation visas ofta bara ett blanksteg ” ” i många IT-system istället för tidsavskiljaren ”T”. Notera också att många IT-system inte klarar av att hantera ISO-standarden, trots att den kom för över tjugo år sedan. Framför allt brukar det vara problem vid hanteringen av tidsavskiljaren och tidszonbeskrivningen.

4.2.2 Fysiska personers namn

Presenteras i formen ”Efternamn, Förnamn”. Tilltalsnamn kan markeras med versaler.

Exempel

”Larsson-Cederstolpe, Eva MARIA Agnes”.

Efternamn som inleds med af, de la, de, di, la, le, van, von eller andra namnpartiklar som är skilda från huvudnamnet ordnas alfabetiskt vanligen efter själva huvudordet i namnet, det vill säga man bortser från partikeln. Normalt skriver man ut partikeln före huvudordet i namnet.

Exempel på sortering

”Andersson, SOFIA Lisa”

”von Atterberg, Beatrice Albertine EFWA Elena”

”Atterdag, WILHELM George Walter”

4.2.3 Personnummer

Personnummer lagras i formen YYYYMMDDXXXX (12 siffror) utan bindestreck enligt Skatteverket, som hanterar folkbokföringen. (Med bindestreck mellan datum och slutsiffror förekommer dock i många organisationer utanför Skatteverket.)

Ex. ”201212121212”.

Presentationen av personnummer sker ofta med ett bindestreck mellan datum och slutsiffror.

Ex. ”20121212-1212”.

4.2.4 Samordningsnummer

Samordningsnummer tilldelas i stället för personnummer till personer som inte är folkbokförda. Bli person senare folkbokförd erhåller den ett personnummer. Samordningsnumret är uppbyggt som ett personnummer fast med talet 60 adderat till födelsedagen. Samordningsnummer lagras i formen NNNNNNNNNNNN (12 siffror) enligt Skatteverkets regler.

4.2.5 Organisationsnummer

Organisationsnummer lagras i formen NNNNNNMMMMMM (12 siffror). Oftast utan bindestreck "-", men ibland med.

4.2.6 Postadress

Uppbyggnaden av svenska postadresser beskrivs av SS 61 34 01 "Posttjänster – Svenska postadresser". Notera att adressat och landnamn ingår egentligen inte i standarden SS 61 34 01.

Vanlig form är:

(adressat)

c/o-namn

utdelningsadress 1

ev. utdelningsadress 2

postnummer postort

(landnamn)

Exempel:

Eva Jönsson

c/o Petterson

Postvägen 12

11111 Poststad

Sverige

4.2.7 Postnummer

Svenska postnummer är i formen NNNNN (00000-99999).

4.2.8 Geografiska koordinater

Hur geografiska koordinater ska anges beskrivs bland annat av ISO 6709. Enligt denna standard anges latitud och longitud i olika kombinationer av grader, minuter och sekunder. Även höjd kan läggas till den horisontella positionen.

Exempel enligt ISO 6709 med bara grader och höjd i meter:

Eiffeltornet: "+48.8577+002.295/", Mount Everest: "+27.5916+086.5640+8850/".

Systemet WGS 84 (World Geodetic System 1984) har tagits fram i USA och används för GPS-positionering, webbaserade kartor, sjökort med mera. I WGS 84 anges latitud och longitud i någon av kombinationerna 1) grader, minuter och sekunder, 2) grader och

decimalminuter eller 3) i decimalgrader. Svenska sjökort har normalt koordinater och rutnät enligt WGS 84 med grader och decimalminuter som enheter. Nordlig latitud är positiv (+, N), sydlig latitud negativ (-, S). Östlig longitud är positiv (+, O, E), västlig longitud negativ (-, V, W). Referenssystemet SWEREF99 (SWEdish REference Frame 1999) baseras på WGS84 och avvikelserna mellan dessa system rör sig oftast endast om några decimeter. Positionering med enbart GPS enligt WGS84 brukar ge en noggrannhet på omkring 10 meter.

Exempel enligt WGS 84 med grader, minuter och sekunder:

Lat 59° 19' 37" N, Long 18° 4' 18" O (Stockholms slott)

Exempel enligt WGS 84 med grader och decimalminuter:

Lat 59° 19,60' N, Long 18° 04,30' E (Stockholms slott)

Tills relativt nyligen användes i Sverige ett referenssystem kallat RT90, ”Rikets Triangelnät”(”Swedish grid”) för den svenska landytan. RT90 är ett rikstäckande rätvinkligt plant koordinatnät som innefattar den svenska landytan. Enligt RT90 anges en position i en x- och y-koordinat. Antal medtagna siffror anger noggrannheten.

Exempel enligt RT90:

X=7087271 Y=1719456 (x- och y-koordinat för centrum i Umeå med en meters noggrannhet.)

Sedan år 2007 använder Lantmäteriverket istället för RT90, projektionen SWEREF99 TM. De allmänna svenska kartorna ritas därför numera utifrån SWEREF99 TM.

Exempel enligt SWEREF99 TM:

N=7088086 E=758900 (x- och y-koordinat för centrum i Umeå med en meters noggrannhet.)

4.2.9 Språkkoder

ISO 639 Two-Letter Language Codes används för att markera språk. Svenska har koden ”sv” (enligt ISO 639-1 med två teckens längd) och ”swe” (enligt ISO 639-2 med tre teckens längd).

4.2.10 Landskoder och svenska länskoder

ISO 3166 Two-Letter Country Codes används för landskoder. Sverige har koden ”SE” enligt ISO-3166. Varje län i Sverige har sedan en vidare uppdelning. Exempelvis ”SE-K”/”SE-10” som betecknar Blekinge län.

4.2.11 Valutakoder

Valutakoder beskrivs av ISO 4217. Svenska kronor har koden ”SEK”, Euro har koden ”EUR”.

4.2.12 Bankkontorsnummer

ISO 9362 ”Banking - Banking telecommunication messages - Bank identifier codes” beskriver formatet för Bank Identifier Codes (BIC), även kallat SWIFT-koder efter förvaltningsorganisationen SWIFT. En BIC-kod är åtta till elva tecken lång.

Ex. ”DEUTDEFF” (Deutsche Bank, Tyskland, huvudkontoret i Frankfurt).

4.2.13 Bankkontonummer

ISO 13616 International Bank Account Number (IBAN) beskriver strukturen för bankkontonummer. IBAN-format för Sverige är:

SEkk BBBB AAAA AAAA AAAA AAAA

där B står för bankkontor/clearingnr och A för kontonummer. SE är landskoden för Sverige enligt ISO 3166, och kk står för kontrollsiffror som räknas fram enligt ”ISO 7064 modula 97-10”. Kontrollsiffrorna gör att man kan validera att överfört bankkontonummer är komplett.

4.2.14 ArkivserieID för punktnotation

SerieID brukar byggas upp av siffror och punkt (”.”). Ibland kan andra skiljetecken förekomma såsom ”-” och ”/”.

Ex. ”1.23.45.3”.

4.2.15 Fil- och katalognamn

Använd inte ”åäöÅÄÖ” i filnamn eller katalognamn.

Ersätt ”mellanslag” med ”underline”, ex. ”a_a.txt”.

Ersätt ”punkt” med ”underline” eller ”p”, ex ”version1_1.txt”, version1p1.txt”.

Undvik helst specialtecken som exempelvis ?*+%&#”~.,:;

Notera att det finns operativsystem som inte kan skilja mellan små och stora tecken. Alla operativsystem och mediaformat sätter dessutom begränsningar för namnlängder som kan skifta avsevärt. Exempelvis klarade de första Microsoft DOS-versionerna bara 8 tecken för filnamn och tre tecken för filändelse.

5. Teckentabeller

5.1 Allmänt

Olika former av skriftsystem har funnits i över fem tusen år hos mänskligheten. Sumererna brukar räknas som den folkgruppering som först skapade olika former av tecken för att symbolisera antal, föremål, händelser, ord med mera. Dessa tecken går fortfarande att läsa okulärt även om betydelsen och uttal oftast inte är helt klar. Detta kan exempelvis jämföras med Harappa-kulturens korta inskrifter som idag är till stora delar oförståeliga.

I västerlandet skapades den första hålkortsstyrda maskinen i början av 1800-talet och hade en primitiv form av avkodningstabell för att styra vävningen av mönster till tyger. Under 1900-talet har sedan konstruerats ett stort antal olika avkodningsschemata som används inom kryptografin och inom IT. De flesta teckentabeller som skapades innan 1990-talet var starkt beroende till leverantör, såsom exempelvis EBCDIC som användes av stordatorer från IBM, MacRoman som skapades av Apple, och de gamla PC DOS-teckentabellerna från Microsoft/IBM.

I detta kapitel berörs vilka teckentabeller som man bör använda för långtidslagring. Märk att i praktiken kan stödet för ISO-8859-1, UTF-8 och så vidare variera mellan olika IT-system och programvaror. En programvara kan uppge att den hanterar informationen efter en specifik teckentabell, när den i praktiken använder en helt annan teckentabell. Detta gäller framför allt äldre programvaror. Det finns bara ett enda sätt om man verkligen vill vara säker att rätt teckentabell används, och det är att testa i praktiken. Vanliga avkodningsfel gäller exempelvis "ääöÅÄÖ", vilka kan generera alla möjliga specialtecken om man har otur. Problemen kan uppstå med alla typer av teckentabeller och beror på att en del programleverantörer inte följt fastställda teckentabell-standarder.

5.2 Rekommendationer

5.2.1 Western Latin-1 (ISO-8859-1)

Teckentabellen ”Western Latin-1” (ISO-8859-1) fungerar oftast vid långtidslagring av texter för majoriteten av västeuropiska språk. I många fall kan man använda Windows-1252 (även kallad ”Windows ANSI”) som ersättare till ISO-8859-1. Detta eftersom Windows-1252 till stora delar är identiskt med ISO-8859-1. (Windows-1252 ersätter ISO-8859-1:s kontrollkoder mellan 0x80 till 0x9F med specialtecken som ”€” och ”%”). Notera att exempelvis samiska inte kan avkodas i sin helhet med ISO-8859-1. Om man tror att man kommer att hantera, eller har krav på att hantera, andra språk än de större västeuropeiska språken, bör man använda Unicode!

5.2.2 UTF-8 och UTF-16 (Unicode Transformation Formats)

Unicode är en standard för att koda alla jordens alfabet och många andra symboler. Unicode kan komprimeras med olika algoritmer, varav UTF-8 (8-32 bitars komprimering) och UTF-16 (16/32-bitars komprimering) är vanligast. UTF-8 och UTF-16 har tagits fram utifrån ISO/IEC 10646: the Universal Character Set.

UTF-8 och UTF-16 kan användas för långtidslagring, även om UTF-8 antagligen är att föredra. Unicode är speciellt lämpligt för att hantera kinesiska, japanska, ryska, sumeriska och andra språk som inte kan hanteras i ISO-8859-1. Teckenstödet för samiska, finska, franska och holländska är också mer heltäckande med Unicode än i ISO-8859-1.

I början av Unicode-filer kan det finnas några kontrolltecken kallade BOM ("Byte Order Mark"). Dessa används för att tolka i vilken ordning och vilken form av Unicode-avkodning som används. Användandet av BOM kan dock ställa till det, eftersom olika leverantörers produkter kan hantera och tolka dem lite olika. I värsta fall kan exempelvis en ISO-8859-1-fil få ett BOM tillagt i början, vilken gör att ISO-8859-1-filen felaktigt tolkas som en Unicode-fil. Med andra ord varningsflagg för felaktigt användande av BOM!¹

ISO-8859-1-filer ska aldrig ha en BOM. BOM bör undvikas för UTF-8 i de flesta fall, eftersom det kan skapa problem med avkodning i olika produkter. BOM behövs däremot för UTF-16.

1 Om man i editor ser att första tecknen är "ï»¿" (hexadecimalt EF BB BF) vet man att en BOM är tillagd för UTF-8. Om sekvensen "þÿ" (hexadecimalt FE FF) eller "ÿþ" (hexadecimalt FF FE) förekommer först i filen vet man att det är en BOM för big-endian UTF-16 respektive little-endian UTF-16 som är tillagd. Detta gäller för Windows ANSI-miljö. I Linux och Mac blir de tecken man ser annorlunda, men de har samma bakomliggande hexadecimala värden.

6. Filformat

6.1 Allmänt

Mängden av filformat är mycket stor och blir ständigt större. I tabell 1 ses några vanligare filformat för långtidslagring och till vilka användningsområde dessa kan korreleras. I avsnitten nedan följer sedan en mer detaljerad genomgång av olika filformat. Märk att detta är rekommendationer, som garanterat kommer att verka omoderna om tjugo år när nya format och ny teknologi tagits fram. Alla filformat kan sägas ha en form av halveringstid, det vill säga den tid tills de börjat ”sönderfalla”, genom att de endast går att läsa till viss del, eller inte alls.

	Kontor*	Scanning	Foto	Ljud	Film	CAD	GIS
PDF/A+JPEG	+	+	?	NA	NA	?	?
JPEG	→	+	+	NA	NA	→	→
PNG	→	?	?	NA	NA	→	→
SVG	→	?	?	NA	NA	? (2D)	?
WebCGM	→	→	→	NA	NA	+(2D)	?
PDF/E+U3D	→	?	→	NA	NA	+(2/3D)	?
GML	NA	NA	NA	NA	NA	NA	+
LPCM WAVE	NA	NA	NA	+	NA	NA	NA
MPEG	NA	NA	NA	+	+	NA	NA
DV DIF	NA	NA	NA	→	?	NA	NA

Tabell 1. Förenklad jämförelse mellan några vanligare långtidslagringsformat och deras respektive användningsområde.

Förklaring: Grönt + = Lämpligt. Gult ? = Kan eventuellt användas. Rött → = Direkt olämpligt, NA = inte tillämpligt, 2D - två dimensioner, 3D - tre dimensioner.

*Med kontor avses format för ordbehandling, kalkylark, epost, OH-presentation, artiklar och liknande.

En generell rekommendation är att information som ska långtidslagras inte bör vara krypterad eller komprimerad. Detta för att öka sannolikheten att informationen kommer att vara tillgänglig på lång sikt framåt. Om man krypterar information ökar risken att den inte går att läsa eftersom det kräver att man bevarar metadata såsom använd kryptoalgoritm, kryptonyckel och i många fall programvara. Om man vill begränsa tillgängligheten ur ett arkivperspektiv är det bättre att använda behörighetskontroller och att säkra de arbetsrutiner som informationen hanteras med. Även yttre komprimering bör undvikas om möjligt eftersom detta också kräver att man vet vilken komprimeringsalgoritm som använts och behöver programvara som kan dekomprimera informationen enligt ursprungligen använd komprimeringsalgoritm.²

2 Detta gäller förstås inte den inre komprimeringen som görs i många filformat såsom JPEG, utan det är den yttre komprimeringen av hela filen eller en samling av filer som avses och som inte ingår i filformatet.



Titel: Rekommendation – Format för långtidslagring

Datum: 2010-03-11

Version 1.0

Något man också kan tänka på är att försöka begränsa antalet filformat och versioner av filformat som man hanterar inom en organisation, eftersom detta underlättar arbetet med framtida formatkonverteringar. Något som också är bra är om man har en förvaltningorganisation för all information som hanteras inom organisationen, inklusive en plan för formatkonverteringar och andra kvalitetsbevarande åtgärder. Riksarkivet behandlar delar av detta område i kapitel 3 "Strategi och planering för bevarande" i RA-FS 2009:1". Se även "TAM 2:2008 REKOMMENDATION – Informationsförvaltning".

6.2 Rekommendationer

6.2.1 W3C XML och ISO 8879:1986 SGML

ISO 8879:1986 SGML (Standard Generalized Markup Language) är en standard för att märka upp texter/kod. SGML baseras ursprungligen på märkspråket GML (Generalized Markup Language) som togs fram vid IBM i slutet av 1960-talet av Charles Goldfarb, Raymond Lorie och Edward Mosher. SGML fick ett stort genomslag vid framtagande av teknisk dokumentation och för skapande av webbsidor under 1980- och 1990-talen, se även avsnitt nedan om webbformat. I SGML beskrivs grammatiken ("strukturen") för en viss informationsmängd med hjälp av en DTD (Document Type Definition).

I slutet av 1990-talet gjordes en förenkling av SGML, som kom att kallas XML (Extensible Markup Language). XML har sedan dess blivit så allmänt använt att det finns som en naturlig del i de flesta former av IT-användningar. Även i XML kan grammatiken ("strukturen") för en viss informationsmängd beskrivas med en DTD på samma sätt som med SGML, eller så kan man använda någon form av så kallat schemaspråk vilket ger en mer exakt grammatik. Exempel på vanliga schemaspråk är W3C Schema och Relax NG. Grammatiken som uttrycks i en DTD / ett schema används för att validera strukturen i en uppsättning XML-dokument. Det finns en klar tendens inom IT-världen att W3C Schema börjar användas allt mer jämfört med DTD. ISO 19757-2 Relax NG (REGular LAnguage for XML Next Generation) med flera andra schemaspråk används ibland som ett alternativ till W3C Schema. Relax NG brukar av sina anhängare betecknas som mer genomtänkt än W3C Schema. W3C Schema har å andra sidan antagligen ett större stöd hos de större leverantörerna.

XML har med åren kommit att utvecklas till en hel familj av standarder som behandlar formatering, transformering, presentation, databasfrågor, signering, kryptering, integrering, grafik med mera.³ XML-teknologi är numera en självklar del i de flesta former av modern kommunikation och långtidslagring. XML är ett mycket brett område och för mer information om XML hänvisas till den mycket omfattande litteraturen för området. En kort introduktion till XML och XML-familjen kan exempelvis fås via wikipedia.

I detta dokument är uppskattningsvis omkring hälften av de rekommenderade filformaten byggda på XML-teknologi, där man exempelvis använder en specifik DTD eller ett schemaspråk för att validera strukturen.

Vanliga filändelser är: .dtd, .ent, .xml, .xsl, .xslt, .xsd, .rng.

MIME-typer: application/xml, application/xslt+xml, text/xml.

3 Exempel på familjemedlemmar är XQuery, XSLT, XSL-FO, Xpath, XML Namespaces, XML Signature, XML Encryption och så vidare.

Rekommendation: En generell rekommendation vid långtidslagring av XML-dokument är att man bör helst lagra det schema man använt för eventuell validering. Detta gäller speciellt om man tvingas att använda ett format som inte är allmänt vedertaget och stabilt. Helst bör man alltid följa standardiserade och öppna format för att underlätta framtida långtidslagring.

För att beskriva grammatiken för XML-dokument bör man i första hand använda W3C XML Schema. I andra hand kan man använda Relax NG eller DTD. XSLT är antagligen tillräckligt stabilt för att användas vid långtidslagring. Stödet för XSL-FO är än så länge mest begränsat till serversidan, och det är tveksamt om det fungerar för långtidslagring.

Observera att XML måste användas med förnuft vid långtidslagring. Det räcker inte med att bara lagra i XML. XML-strukturen som används måste dessutom vara dokumenterad, genomtänkt, stabil, allmänt använd och allmänt tillgänglig. Man kan göra fullkomligt vansinniga saker med XML om man inte vet vad man håller på med.

6.2.2 PDF/A-1 (ISO 19005)

PDF/A-1 (PDF-Archive) kan i grova drag beskrivas som en förenkling utifrån PDF Version 1.4 (Portable Document Format). PDF/A-1-formatet stödjer inte kryptering, komprimering, inbäddade filer, referenser till externa filer, multimedia, operationer i formulärfält eller transparens hos bilder. Använda typsnitt ska vara inbäddade och färgrymder (colour spaces) ska vara i applikationsoberoende format. Inbäddade digitala signaturer tillåts. Inbäddat metadata krävs till viss del. Ytterligare metadata kan också tillföras vid behov för att få dokument självförklarande. PDF/A-1 finns för två nivåer PDF/A-1b och PDF/A-1a. Det senare alternativet ställer krav på att man måste tillföra mer metadata om textens struktur, innehåll och språk. Nivån PDF/A-1b ingår som en del i nivån PDF/A-1a.⁴

Rekommenderade filändelser är: .pdf.

MIME-typ: application/pdf eller text/pdf.

Rekommendation: PDF/A-1 är ett förstahandsval för att långtidslagra alla former av kontorsdokument, scannande dokument, rapportutskriften et.c. Om man är tveksam i vilket filformat man ska långtidslagra i, så brukar PDF/A-1 vara ett "standard-tips". Nivån PDF/A-1b brukar vara tillräcklig för de vanligaste arkivsammanhangen. Den utvidgade nivån PDF/A-1a behöver oftast bara användas om man har speciella krav på att bevara exempelvis hur ett dokument är uppbyggt. Notera att bara för att ett dokument är lagrat i PDF så räcker inte detta för långtidslagring, utan det måste vara i PDF/A-1-format.

4 PDF/A-1b är en äkta delmängd till PDF/A-1a.

6.2.3 ODF (ISO 26300)

ODF (Open Document Format for Office Applications), även kallad OpenDocument är en uppsättning XML-baserade format för att lagra och utbyta vanliga kontorsdokument (ordbehandling, kalkylblad, grafik, presentation och liknande) mellan olika program. Notera att ODF inget har att göra med Microsofts format OOXML.

Rekommenderade filändelserna är .odt, .ods, .odp, .odg och .odf.

MIME-typer: application/vnd.oasis.opendocument.text,

application/vnd.oasis.opendocument.presentation,

application/vnd.oasis.opendocument.spreadsheet och

application/vnd.oasis.opendocument.graphics.

Rekommendation: Använd PDF/A-1 i första hand för lagring av kontorsdokument. I dagsläget är det osäkert om ODF är tillräckligt spritt för att fungera för långtidslagring. Det är möjligt ODF kommer att få en tillräcklig bedömd livslängd och stabilitet om några år för att fungera för långtidslagring, beroende på vilket stöd ODF får hos användarna. Ett möjligt alternativ är att man parallelllagrar i både ODF och PDF/A-1. På detta sätt får man dels en möjlig arbetsversion som man lättare kan editera och dels en mer statisk arkivversion. Att parallellagra i både ODF och PDF/A-1 brukar inte göra så stor skillnad minnesmässigt, jämfört med att endast lagra i PDF/A-1. Detta eftersom en ODF-fil brukar uppta en bråkdel av en PDF/A-1-fils storlek.

(För jämförelse kan nämnas att Microsofts format OOXML bedöms idag som ännu mer instabilt och komplext än ODF och därför direkt olämpligt för långtidslagring. Se även avsnittet om leverantörsberoende kontorsformat.)

6.2.4 Leverantörsberoende kontorsformat

Det finns en uppsjö av olika leverantörsberoende filformat för ordbehandling, kalkylering, grafik, layout och presentation. Leverantörsberoende format kan vara icke-standardiserade, ”de facto”-standardiserade eller till och med ”de jure”-standardiserade. Det räcker tyvärr inte med att ett format är fastställt av ett internationellt standardorgan för att det ska vara lämpligt för långtidslagring.

Rekommendation. Man använder aldrig leverantörsberoende format för långtidslagring om man kan undvika det, eftersom de ofta är för instabila, för ostrukturerade och med beroenden till olika leverantörers affärsstrategier. Detta ger svåruppskattade livslängder och därmed blir formaten olämpliga för långtidslagring. Notera dock att leverantörsformat fungerar utmärkt för kortsiktig lagring (omkring 1-5 år).

6.2.5 email-format

Det finns ett antal produktberoende filformat för paketering och lagring av email. Inget av de produktberoende email-filformaten fungerar bra för långtidslagring. Se även avsnittet om leverantörsberoende kontorsformat.

Rekommendation: email som ska långtidslagras konverteras till 1) oformaterat textformat, 2) PDF/A-1, 3) XHTML eller 4) XML-uppmärkt struktur med CSS. Ev. bilagda filer måste även de konverteras till format som klarar långtidslagring.

6.2.6 Webbformat (XML, XHTML, HTML, CSS, XSLT, XSL-FO...)

Grovt sätt finns det tre alternativ för långtidslagring av webbplatser; 1) komplett Configuration Management av hela webbplatsen inklusive databaser, 2) transaktionslagring som lagrar alla utföra transaktioner eller 3) regelbunden ”nedfrysning”, med eller utan bakomliggande databaser, om man vill se hur en webbplats såg ut vid ett specifikt tillfälle. Alternativ 1) eller 2) brukar användas om man har höga krav på bevisbarhet för vilken information som gjordes tillgänglig vid olika tillfällen. Alternativ 3) är vanligast och brukar användas för att spara olika versioner av samma webbplats vid olika ögonblick. Om man använder alternativ 3) brukar insamlingen ske genom 3a) nedtankning via www med ett webbcrawlingsverktyg, 3b) komplett kopiering av webbplatsens interna lagringsstrukturer och innehåll, eller 3c) regelbunden export ur bakomliggande webbsystem och databaser till långtidslagringsformat med metadata-beskrivningar. Oberoende av vilket tillvägagångssätt man väljer så har man nästan alltid problemet med att många webbformat är instabila över tiden. De har kort ”halveringstid” innan de sönderfaller.

Grundformaten för World Wide Web är HTML (Hyper Text Markup Language) och XHTML (Extensible HTML). Fysikern Tim Berners-Lee vid CERN tog år 1990 fram första versionen av HTML. I kombination med andra teknologier såsom TCP/IP och HTTP följde därefter en explosionsartad informationsutveckling, där hemsidor började skapas i alla världens hörn från Chile till Tibet. XHTML är en vidareutveckling av HTML. Den stora skillnaden är att XHTML följer XML-standarden, vilket bland annat gör att dokument är lättare att bearbeta, och läsa av maskinellt. HTML är byggd på SGML (Standard Generalized Markup Language, ISO 8879:1986).⁵ XHTML eller HTML brukar kombineras med CSS eller XSL-FO för att styra presentationen av hemsidor. XSLT används för att transformera om XML-filer till i princip vilket filformat eller vilken struktur som helst. XSLT och XSL-FO är betydligt mer kraftfullt än CSS, men används oftast bara på serversidan. CSS används på klientsidan.

Några typiska filändelser är .xml, .xsl, .css, .xhtml, .htm, .html, .xht, .xsl, .xslt.

Några vanliga MIME-typer: application/xml, text/xml, application/xhtml+xml, text/xhtml+xml, text/css, application/css-stylesheet, application/xslt+xml.

5 För att göra saken mer förvirrad arbetar W3C sedan år 2004 med att slå ihop XHTML 1.1 och HTML 4.01 till en enda standard med arbetsnamnet ”HTML5”.

Rekommendation: Det finns många olika filformat som används när man skapar statiska hemsidor eller genererar dynamiska hemsidor. Bara ett fåtal av dessa fungerar väl för långtidslagring. De format som fungerar bäst för långtidslagring av hemsidor är XML, XHTML Strict, CSS och XSLT. I andra hand kan man använda HTML 4.01 istället för XHTML.

Det finns ingen möjlighet att på sikt garantera funktionaliteten hos dagens script, serverbaserad kod och liknande. Detta på grund av den snabba tekniska utvecklingen vilken gör äldre teknologier obsolet. Det som är modernt idag är garanterat "ute" om tio år. Helst ska därför all form av exekverbar kod, förutom XSLT, tas bort och sidor omvandlas till statiska sidor för att sidor lättare ska kunna långtidlagras. Å andra sidan så går det i många fall inte att ta bort scriptkod på klientsidan utan att få olika former av förluster i funktion och innehåll. Detta eftersom många webbplatser har konstruerats för att vara beroende av scriptkod på klientsidan för att fungera. Om man därför vet vid framtagandet av en ny webbplats att man vill kunna långtidslagra den så bör man undvika klientscript helt, eller i andra hand använda rent ECMA-script (ECMA 262) i så hög grad som möjligt, utan de leverantörsspecifika tillägg som finns i Jscript/JavaScript.

Länkar mellan sidor som ska långtidslagras måste vara sins emellan relativa⁶. Eventuella absoluta⁷ länkar måste omvandlas till relativa. Länkar till externa resurser eller databaser går inte heller att garantera över tiden, eftersom URL:ar är tillfälliga. Därför måste man överväga om man ska 1) lyfta in en kopia av den resurs som sidan länken refererar till, 2) acceptera att länken snabbt kan bli inaktuell eller 3) avaktivera länken men ha kvar länktextern till den externa sidan.

Det finns ett antal organisationer som forskar runt långtidslagring av webbplatser, exempelvis Library of Congress, National Library of New Zealand, LDB-centrum och Kungliga Biblioteket i Sverige. TAM-Arkiv planerar att i framtiden i samarbete med andra organisationer studera vidare frågan runt långtidslagring av webbplatser.

6.2.7 JPEG (ISO 10918)

JPEG (Joint Photographic Expert Group) är egentligen endast en komprimeringsalgoritm. JPEG-komprimering kan lagras i olika filformat såsom PDF/A, "JPEG Interchange Format" och JFIF (JPEG File Interchange Format). JPEG komprimering kan vara destruktiv (lossy) eller bevarande (lossless). Lossy-komprimering sker med algoritmen DCT (Discrete Cosine Transform). Få applikationer har stöd för lossless-komprimering enligt JPEG-standarderna. Notera att varje ny lossy JPEG-komprimering ger en ny informationsförlust.

Vanliga filändelser för JPEG-filer är .jpg, .jpeg, .jpe, .jfi, .jfif, .j, .jif, och .jmh.
MIME-typ: image/jpeg.

6 Ex. "download/rfc-index.txt".

7 Ex. "http://www.ietf.org/download/rfc-index.txt".

Rekommendation: JPEG kan användas för att långtidslagra rasterade bilder och fotografier i gråskala och färg. För långtidslagring används helst låg grad av destruktions (Kvalité 85-95%). Kan användas i kombination med exempelvis MIX.

6.2.8 JPEG 2000 (ISO 15444)

JPEG 2000 är en vidareutveckling av JPEG och använder mer effektiva och komplicerade komprimeringsalgoritmer. Komprimeringen kan vara både lossless och lossy.

Rekommenderade filändelser för JPEG 2000-filer är .jp2 (ISO/IEC 15444-1) och .jpx (ISO/IEC 15444-2).

MIME-typer: image/jp2, image/jpeg2000 och image/jpx.

Rekommendation: Än så länge mindre lämpligt för långtidslagring då formatet är mer komplext och inte har lika utbredd användning jämfört med vanlig JPEG. JPEG2000-bilder har dock den fördelen att de oftast är av bättre kvalitet avseende färgåtergivning och "klarhet" än JPEG-bilder. Kan användas i kombination med exempelvis MIX.

6.2.9 PNG (ISO 15948:2004)

PNG (Portable Network Graphics) är en icke-kommersiell vidareutveckling av det rasterade bildformat GIF. GIF har dock fortfarande större spridning än PNG och stödjer animering, vilket PNG inte gör. PNG är lossless.

Rekommenderad filändelsen är .png.

MIME-typ: image/png.

Rekommendation: Kan används för att långtidslagra rasterade bilder och fotografier. Kan användas i kombination med exempelvis MIX.

6.2.10 GIF 89a (Graphics Interchange Format)

GIF 89a (Graphics Interchange Format) är ett filformat för rasterade bilder. Används framför allt till hemsidor. GIF togs fram av CompuServe/Unisys.

Rekommenderad filändelse är .gif.

MIME-typ: image/gif.

Rekommendation: GIF bör undvikas för långtidslagring därför att det är ett leverantörsbundet format. Använd hellre PNG som är en ersättare till GIF.

6.2.11 TIFF

TIFF (Tagged Image File Format) är ett äldre filformat ursprungligen framtaget av Aldus/Adobe för lagring av rasterade bilder. Formatet innehåller en mycket stor mängd olika profiler och komprimeringsalgoritmer. Exempelvis kan man använda JPEG-algoritmer i TIFF. De flesta program brukar endast stödja delar ur de olika versionerna av TIFF vilket kan ställa till det.

Rekommenderad filändelse är .tif eller .tiff.
MIME-typ: image/tiff.

Rekommendation: Använd hellre PDF/A eller JPEG/JPEG2000. (Om man absolut måste använda TIFF bör man använda TIFF 6.0/Lempel-Ziv-Welch lossless för färg-/gråskala och TIFF 6.0/CCITT T.6 "grupp 4" för svart/vitt.)

6.2.12 SVG (W3C SVG)

SVG (Scalable Vector Graphics) är ett XML-format för lagring av 2-dimensionell grafik, ex. CAD. SVG kan hantera text, rasterad och vektoriserad grafik. Formatet har även stöd för viss animering.

Rekommenderad filändelse är .svg.
MIME-typ: image/svg+xml, image/svg+xml.

Rekommendation: SVG kan användas för långtidslagring av 2-dimensionella vektoriserade ritningar, illustrationer och animationer.

6.2.13 WebCGM och CGM (ISO/IEC 8632:1999)

CGM (Computer Graphics Metafile) är ett format för utbyte av 2-dimensionell vektor- och vektor/raster-grafik. WebCGM är en anpassning av CGM som tagits fram mellan W3C och CGM Open.

Rekommenderad filändelse är .cgm.
MIME-typ: image/cgm.

Rekommendation: WebCGM 2.0 eller senare kan användas för 2-dimensionella tekniska ritningar, tekniska illustrationer och liknande.

6.2.14 U3D (ECMA 363)

U3D (ECMA Universal CAD 3D) är en standard för att kunna lagra och utbyta 3-dimensionella ritningar.

Filändelsen för formatet är: .u3d.

Rekommendation: U3D kan sannolikt användas för att långtidslagra 2- och 3-dimensionella ritningar.

6.2.15 PDF/E (ISO 24517-1:2008)

PDF/E (PDF-Engineering) är i likhet med PDF/A en förenkling och anpassning av PDF, i detta fall utifrån version 1.6. PDF/E-1 är i första tänkt för utbyte av tekniska ritningar och dokumentation mellan olika leverantörers system.

Rekommenderad filändelse är .pdf.

MIME-typ: application/pdf eller text/pdf.

Rekommendation: PDF/E-1 tillsammans med U3D kan sannolikt användas för att långtidslagra teknisk dokumentation och ritningar i 2- och 3-dimensioner.

6.2.16 X3D (ISO 19775 m.fl.)

X3D är en uppsättning nya XML-format för att kommunicera information om 3-dimensionella objekt. X3D är en efterföljare till VRML (Virtual Reality Modeling Language). X3D har omfattande funktionalitet, inklusive stöd för ljud, bild och animationer.

Vanliga filändelser är: .x3dv, .x3d och .x3db.

MIME-typ: model/x3d+xml

Rekommendation: Det är i dagsläget oklart hur stabilt X3D är, då det är ett mycket komplext format som antagligen måste begränsas för att fungera för långtidslagring. Använd hellre andra format för långtidslagring som PDF/E med U3D, SVG eller WebbCGM.

6.2.17 GML (ISO 19136:2007)

GML (Geography Markup Language) är en standard framtagen inom The Open Geospatial Consortium för att kunna utbyta geografiska ritningar och data mellan olika leverantörers GIS-system.

Rekommenderad filändelse är .gml.

Rekommendation: GML är sannolikt tillräckligt stabilt för att kunna användas för att långtidslagra geografiska information.

6.2.18 WAVE LPCM

WAVE är ett ljudfilformat framtaget av IBM och Microsoft. WAVE-formatet tillåter både komprimerad och okomprimerad lagring. European Broadcast Union (EBU) använder en uttökad variant av den okomprimerade profilen WAVE LPCM (Waveform Audio File Format with Linear Pulse Code Modulation bitstream). Nackdelarna med WAVE LPCM-filer är att de tar stort lagringsutrymme och att ljudfilerna i praktiken brukar få bli högst 2 GB. Som ett sidospår kan nämnas att vanliga CD-ljudskivor använder LPCM-algoritm, dock inte WAVE LPCM, för att lagra ljud; i detta fall i två kanaler med 44'100 samplingar/s om 16 bitar.

Vanligaste filändelserna är .wav or .wv.
MIME-typ: audio/wav eller audio/wave.

Rekommendation: WAVE LPCM okomprimerat med minst 96 kHz om 24 bitar bör användas för långtidslagring för ljudinspelningar av hög kvalité, om man vill kunna bearbeta och filtrera ljudet i framtiden i ljudprofessionella sammanhang. (Minst 16 bitar om 48 kHz okomprimerat rekommenderas enligt EBU.)

Som distributions- och konsumentformat används MP3 istället. Nackdelen med MP3 är att detta format har förstörande komprimering, vilket gör att lagrat ljud inte kan bearbetas och behandlas i framtiden utan defekter och förvrängningar.

6.2.19 MPEG-1 (ISO 11172:1993, ITU-T standard H.261) inkl. MP3

MPEG-1 är ett destruktivt format som tagits fram för att lagra komprimerat ljud och bild , motsvarande VHS-kvalité och Video CD. Bildkvalitén är begränsad. I standarden ingår även det mycket populära ljudformatet MP3, eller som det egentligen heter ”MPEG-1 Audio Layer 3”. En MP3-fil vid 128 kbit/s sampling tar omkring en tiondel i lagringsutrymme jämfört med WAVE LPCM.

Vanligaste filändelserna är: .mpeg, mpg och mp3.
MIME-typ: audio/mpeg, video/mpeg.

Rekommendation: Om man vill spara lagringsutrymme, och då bildkvalitén inte är väsentlig kan man använda MPEG-1 för långtidslagring av video. MP3 med en bit-rate om minst 128 kbit/s fungerar utmärkt för långtidslagring av mono- eller stereoljud för den vanlige användaren. MP3 ska dock inte användas att långtidslagra professionellt ljud såsom konsertupptagningar, fyra-kanalers-film ljud, studioupptagningar och liknande ljud som kan behöva bearbetas eller filtreras.

6.2.20 MPEG-2 (ISO 13818:1995)

MPEG-2 är ett konsument- och distributionsformat för film. Formatet är en vidareutveckling av MPEG-1. MPEG-2 är fullt bakåtkompatibelt med MPEG-1, vilket gör att en MPEG-2-spelare även kan läsa av en MPEG-1-fil. MPEG-2-filer tar mer lagringsutrymme än MPEG-1-

filer. MPEG-2 har en högre bildkvalite än MPEG-1, och används för komprimering av video till bland annat DVD-skivor och HDTV. MPEG-2 har ett antal olika profiler med olika upplösning, samplingsfrekvens och bandbredd. Exempelvis upp till en upplösning om 720×576 för DVD, och upp till 1920×1080 för HDTV.

Vanligaste filändelserna är: .mpeg och mpg.

MIME-typ: video/mpeg.

Rekommendation: MPEG-2 fungerar utmärkt att använda för långtidslagring av video/film för den vanlige användaren. I bildprofessionella sammanhang där man inte kan tillåta omfattande destruktiv komprimering eller behöver lagra högkvalitativ film för framtida redigering så räcker inte alltid MPEG-2 med lägre bit-rate till. Vill man ha MPEG-2-filmer av hög kvalitet ("produktionskvalité") ska man endast ha med I-frames⁸ med en hög bit-rate såsom 50 Mbit/s och chroma sampling om 4:2:2, men detta är förstas kraftigt lagringsplatskrävande. Se även avsnittet om DV och DV-DIF.

6.2.21 MPEG-4 (ISO 14496), DivX och Xvid

MPEG-4 bygger på Apples format Quicktime. MPEG-4 är en mycket komplex standard som innehåller ett antal olika algoritmer och metoder för komprimering av ljud och video, hantering av virtuella objekt, copyright, textströmmar et cetera. MPEG-4 är under vidareutveckling. MPEG-4 part 2 Advanced Simple Profile (ASP) används av det proprietära videoformat DivX och det öppna videoformatet Xvid (GNU-licens). En av fördelarna med Xvid och DivX är att filerna generellt tar mindre lagringsutrymme. MPEG-4 part 10 Advanced Video Coding/H.264 används i Blu-ray-skivor.

Vanliga filändelser är: .mp4., .m4a, divx och .xvid.

MIME-typ: video/mp4, audio/mp4 och application/mp4.

Rekommendation: MPEG-4 part 2 (H.263) ASP tillsammans med Xvid, eller möjligen DivX, kan eventuellt användas för långtidslagring av film. Vill man ha MPEG-4-filmer av hög kvalitet ("produktionskvalité") ska man endast ha med I-frames med en hög bit-rate och chroma sampling om 4:2:2. Det är svårt att uppskatta livslängden för filformat byggda på MPEG-4 Part 10/H.264, och detta beror delvis på vilket marknadsgenomslag som Blu-ray-skivor får innan detta format ersätts av ny teknologi. I dagsläget är det osäkert om MPEG-4 är tillräckligt begränsat och stabilt för att fungera för långtidslagring, eller inte. Inom några års sikt kommer antagligen situationen för MPEG-4 att bli tydligare.

8 I-frame ("Intra-coded picture") är en typ av frame som inte är beroende av föregående eller efterkommande frame för avkodning. De andra typerna heter B-frame ("Bi-predictive picture") och P-frame ("Predicted picture").

6.2.22 DV och DV-DIF

DV (Digital Video) innehåller en omfattande familj av olika öppna och leverantörsberoende format för inspelning och uppspelning av digital video. DV utvecklades ursprungligen för digitala tv-/videokameror på 1990-talet. Varje frame komprimeras för sig, vilket gör att det går att lätt redigera filmer bild för bild, och att filmen håller en högre bildkvalité per frame. ”Råformatet” för DV är DV-DIF (Digital Video Digital Interface Format). DVCPRO25 (25 Mbit/s, 4:1:1 bildsamplingsstruktur/chroma sampling) och DVCPRO50 (50 Mbit/s, 4:2:2 bildsamplingsstruktur/chroma sampling) är videofORMAT framtagna av Panasonic som är relativt vanliga.

Vanligaste filändelserna är: .dv och .dif.

MIME-typ: video/x-dv

Rekommendation: I videoprofessionella sammanhang, räcker inte alltid MPEG-standarderna till. Om så är fallet kan man pröva att lagra i en DV-DIF-variant. Exempelvis DVCPRO50 med tillägget SMPTE 314M-1999 Television, 625 linjer, 50 Hz, 50 Mbit/s bit-rate och 4:2:2 chroma sampling, vilket ger en hög bildkvalité och möjlighet att redigera filmen i framtiden. Livslängden för DV-DIF är svår att uppskatta. Ett alternativ till att bara lagra i bara DV-DIF är att parallelllagra i både DV-DIF med en högre bit-rate och i MPEG-2 med en lägre bit-rate. Detta eftersom MPEG-2 antagligen har en längre livslängd, medan DV-DIF ger möjligheterna till redigering och högre bildkvalité, men detta angreppssätt är förstas lagringsplatskrävande.

6.2.23 Vårdinformation (DICOM, HL7, HISA...)

DICOM (Digital Imaging and Communications in Medicine) är en standard för att lagra och överföra medicinska bilder med metadata. DICOM-standarderna är omfattande och innehåller exempelvis krav för fysiskt media, nätverksprotokoll, filformat och abstrakt datamodell. Bilder i DICOM kan komprimeras med ett antal olika algoritmer såsom JPEG lossy och JPEG lossless. Stora delar av DICOM, såsom rekommendationerna för fysiskt media och kommunikationsprotokoll, är direkt irrelevanta för långtidslagring. Endast de delar av DICOM som beskriver fil- och dataformat har betydelse vid långtidslagring, se ”Part 10: Media Storage and File Format for Media Interchange”.

Det europeiska standardorganet CEN arbetar sedan många år med att standardisera utbytet av vårdinformation mellan vårdsystem. Bland annat inom ”EN 13606 - a Standard for EHR System Communication” och ”ENV 12967 - Standard Architecture for Healthcare Information Systems”, det senare arbetet kallas ofta för ”HISA”.

Parallellt med CEN:s arbeten inom vårdinformationsområdet har i USA bedrivits arbeten inom motsvarande område inom standardorganisationen HL7 (Health Level Seven). HL7 CDA (HL7 Clinical Document Architecture) är utvecklat för att kommunicera vårdinformation mellan system i form av XML-dokument.

Det bedrivs i Sverige sedan länge ett sisyfos-arbete runt XML-vokabulärer för kommunikation av vårdinformation mellan system. Detta arbete bedrivs mellan landstingen, delar av

Kommunförbundet, Carelink, SIS (Svenska Industrins Standardiseringskommission) och olika privata aktörer inom vårdsektorn. Exempelvis inom ”RIV - Regelverk för elektroniskt informationsutbyte inom vård och omsorg”. Det är dock svårt att bedöma om resultaten från dessa arbeten är tillräckligt stabila för att fungera även för långtidslagring och inte bara kommunikation i dagsläget.

Vanliga filändelser för DICOM är: .dcm och .dic.

Rekommendation: DICOM kommer sannolikt att fortsätta att användas under lång tid framöver inom sjukvården, vilket gör att filformatet antagligen fungerar att lagra information för åtminstone 10-20 år, så länge som man undviker att komprimera bilder med mer udda algoritmer, använder leverantörsspecifika extensions eller krypterar filer. Notera dock att standarden är komplex och under ständig vidareutveckling, vilket försvårar möjligheten till långtidslagring.

Det finns än så länge ingen allmänt accepterad stabil europeisk eller svensk XML-standard som fungerar för långtidslagring av vårdinformation. I brist på bättre kan man pröva att lagra i enligt med de generella riktlinjerna i 1) HL7 CDA med svenska anpassningar eller 2) HISA med svenska anpassningar. Parallell lagring i PDF/A-1 är också antagligen att rekommendera tills vidare.

6.2.24 Generella format för lagring av tekniskt data (CDF, HDF...)

Det finns en mycket stor mängd filformat för lagring och utbyte av teknisk och naturvetenskaplig information, såsom AML (Astronomical Markup Language), CML (Chemical Markup Language), eXtensible Data Format (XDF) och MathML. I detta avsnitt behandlas bara de av mer generell karaktär. Common Data Format (CDF) är ett format utvecklat av NASA för att lagra och hantera flerdimensionella datamängder. HDF (Hierarchical Data Format) är ett generellt format framtaget av U.S. National Center for Supercomputing Applications (NCSA) för att lagra flerdimensionella mycket stora datamängder i en hierarkisk struktur. HDF kan även i vissa fall användas för att migrera data från en databas till en annan.

Vanliga filändelser är: .h5 och .cdf.

Rekommendation: HDF5 och CDF kan antagligen användas för att lagra data för åtminstone 10 år.

6.2.25 Generella format för lagring av data ur databaser (ADDML, SIARD...)

ADDML (Archival Data Description Markup Language) togs ursprungligen fram av norska Riksarkivet för att beskriva databasers strukturer och innehåll, det vill säga metadata till en arkivleverans av en databas. Svenska Riksarkivet samarbetar numera med norska Riksarkivet

runt ADDML. Bland annat SCB har börjat använda ADDML för att beskriva databasleveranser till Riksarkivet.

SIARD (Software Independent Archiving of Relational Databases) är ett filformat med tillhörande verktyg som ursprungligen togs fram av schweiziska Riksarkivet för långtidslagring av relationsdatabaser. En SIARD-fil byggs upp av en samling av XML-filer som paketerats utan kompression i en ZIP64-container. Till skillnad från ADDML så struktureras både data och metadata med XML i en SIARD-fil, där varje tabell blir en separat XML-fil. SIARD inkluderar även en uppsättning java-baserade verktyg som kan användas för att skapa och tillgängliggöra SIARD-filer. SIARD är det format som används inom EU-projektet ”Planets”.

Inom Apache-projekten så finns ett projekt kallat Torque som arbetar med att hantera data i generella relationsdatabaser. Sveriges Lantbruksuniversitet använder delar ur Apache Torque för att lagra data ur forskningsdatabaser.

Redan år 2000 arbetade en arbetsgrupp⁹ inom ISO fram ett första förslag för export ur relationsdatabaser via JDBC till XML. År 2008 fastställdes den nya ISO SQL-specifikationen (ISO 9075:2008), där kapitel 14 rör omvandlingar mellan SQL och XML.¹⁰ I denna specifikation (ISO 9075-14:2008) berörs mappningar fram och tillbaka mellan SQL och XML, inklusive hanteringen av NULL-värden vilket kan ställa till det vid ofullständig omvandling från SQL till XML. Notera gärna att synsätten som förs fram i denna ISO-standard i en del fall inte följer de som brukar rekommenderas i XML-kretsar, vad gäller namngivning av element och så vidare.

Rekommendation: Om man vill långtidslagra relationsdatabaser bör man i första hand använda SIARD eller ADDML. Man kan i vissa fall få vägledning för omvandlingar från relationsdatabaser till XML utifrån ISO 9075-14:2008. Även HDF kan antagligen i vissa fall vara intressant att undersöka. Ytterligare ett alternativ är att denormalisera en databas till ett eller flera dokumentträd i XML-struktur. I detta fall bör man endast följa ISO 9075-14:2008 i tillämpliga delar såsom för datatyper. Exempelvis kan denormaliserad lagring vara aktuell för diaries och andra former av enklare postorienterade register. TAM-Arkiv planerar att arbeta vidare med detta område och återkomma med mer detaljerade rekommendationer i samarbete med andra arkivinstitutioner.

6.2.26 Format för finans och redovisning (SveFaktura, XBRL...)

Det finns ett antal internationella format för finans och redovisning såsom OAGIS, ebXML, XBRL, FpML med flera. Dessa format och ramverk är i dagsläget fokuserade på informationsutbyte och inte på lagring, även om detta kan ändras i framtiden. Ramverk som OAGIS och ebXML är dessutom mycket komplexa och måste anpassas och skalas ner för att användas i en specifik situation.

⁹ ISO/IEC JTC 1/32: Data Management and Interchange/WG3: Database Languages.

¹⁰ ISO/IEC 9075-14:2008 Information technology -- Database languages -- SQL -- Part 14: XML-Related Specifications (SQL/XML).

I Sverige har ESV/e-nämnden utifrån ebXML-ramverket tagit fram ett XML-vokabulär för fakturor kallat SveFaktura. Svenska SIE-gruppen har tagit fram ett annat XML-vokabulär för utbyte av redovisningsinformation mellan system kallat SIE-XML. Tyvärr förvaltas dock detta vokabulär inte längre av SIE-gruppen eller någon annan organisation. Svenska Riksarkivet har skapat ett vokabulär kallat BTLA (BalancesTransactionsLedgerentriesAssets) utifrån SIE-XML för att kunna ta emot exempelvis transaktionslistor.

Rekommendation: SveFaktura kan användas för att långtidslagra fakturor, om man skulle vilja det. Dock brukar fakturor oftast inte behöva lagras längre än tio år, så nyttan med långtidslagring med SveFaktura är tveksam. För ekonomiska rapporter, ex. kvartals- och årsrapporter, kan möjligen XBRL användas för långtidslagring. Om man behöver långtidslagra annan form av ekonomisk information i strukturerad form kan Riksarkivets BTLA vara värt att pröva. Annars är som vanligt PDF/A-1 att rekommendera.

6.2.27 Arkivstandarder (EAD, EAC-CPF, PREMIS, METS, MIX, textMD...)

Encoded Archival Description (EAD) används för att lagra och utbyta arkivförteckningsinformation. EAD är bra på att beskriva arkivmaterial lagrade på traditionella media såsom papper, COM (Computer Output Microfilm) och fotopapper. EAD har även visst enklare stöd för förteckning av digitala resurser såsom databaser.¹¹ EAC-CPF (Encoded Archival Context - Corporate Bodies, Persons, and Families) används för att utbyta information om arkivbildare, det vill säga de organisationer, konstellationer eller personer som skapat ett godtyckligt arkiv. EAC-CPF revision 2010 har ersatt den tidigare beta-versionen av EAC. EAC används exempelvis av Riksarkivets NAD-grupp för att ta emot arkivbildarbeskrivningar.

Library of Congress förvaltar ett antal olika XML-baserade arkivstandarder, förutom samförvaltningen av EAD och EAC,¹² såsom PREMIS, METS, textMD och MIX. PREMIS (PREservation Metadata: Implementation Strategies) kan användas för att lagra metadata om informationsobjekt som ska långtidslagras. METS (Metadata Encoding and Transportation Standard) används för paketering av information som ska långtidslagras. METS kan också användas för att beskriva den inre strukturen för informationen som ska långtidslagras. MIX (NISO Metadata for Images in XML) kan användas för att beskriva tekniskt metadata för ett fotografi eller en bild. TextMD (Technical metadata for Text) har en motsvarande funktion som MIX fast för textbaserade dokument istället för fotografier..

Ofta kombineras flera arkivstandarder med varandra och med andra generella metadatastandarder. På detta sätt kan man beskriva den information man långtidslagar på flera nivåer, såsom kontextuell nivå, arkivbildarnivå, arkivnivå, aggregatnivå och instansnivå. Dessa beskrivningar används sedan för att tolka och förvalta informationen som lagras.

11 Detta kan göras med elementen <dao>, <daodesc>, <daogrp> och <daoloc>.

12 EAD samförvaltas mellan Society of American Archivists och Library of Congress. EAC-CPF samförvaltas av Society of American Archivists och Staatsbibliothek zu Berlin.

Rekommendation: Senare versioner av EAC, EAD, PREMIS och METS kan sannolikt användas tillsammans eller separat för att hantera digitala leveranser och för långtidslagring av digital information. MIX och textMD torde också kunna gå att använda för långtidslagring. Märk att senare versioner av exempelvis EAC och METS inte är fullt bakåtkompatibla med tidigare versioner. TAM-Arkiv planerar att i framtiden ge mer detaljerade anvisningar för vilka arkivstandarder som kan användas för digitala leveranser till TAM-Arkiv.

6.2.28 Generella metadatastandarder (Dublin Core, RDF, PRISM...)

Det finns en uppsjö av standarder och standardiseringsförsök för generellt metadata. W3C Resource Description Framework (RDF) används för att beskriva relationer mellan olika objekt. Dublin Core (ISO 15836:2003) framtaget inom Dublin Core Metadata Initiative är en mycket vanlig standard för att mycket övergripande beskriva dokument, hemsidor, fotografier och liknande. Dublin Core är framtaget för biblioteksvärldens behov med de styrkor och begränsningar detta medför. The Publishing Requirements for Industry Standard Metadata (PRISM) inkluderar Dublin Core och RDF och ger dessutom möjlighet att lägga till ytterligare metadata såsom rättigheter, publicering och liknande. PRISM används bland annat inom förlagsvärlden. "ISO 23081-1:2004 Records management processes - Metadata for records" är en ISO-standard för metadatabeskrivning av dokument. Det finns många andra standarder med större eller mindre genomslag som används för att beskriva dokument med metadata; exempelvis UK e-Government Metadata Standard (inom UK eGov/GovTalk), Metadata Object Description Schema (MODS), Victorian Electronic Records Strategy (VERS), Australian Government Recordkeeping Metadata Standard och Model Requirements for the Management of Electronic Records (MOREQ).

Rekommendation: Det går inte att enkelt ge en allmän rekommendation för vilken generell metadatastandard som man ska använda för långtidslagring, eftersom det beror på verksamhetens egna behov av detaljgrad och typ av information som man vill beskriva. Varje generell metadatastandard har sina styrkor och svagheter. RDF och Dublin Core, enligt den XML-standard som DCMI, anvisar fungerar utmärkt för långtidslagring. Generella metadatastandarder kombineras oftast med arkivstandarder för att få en fungerande långtidslagring. TAM-Arkiv planerar att i framtiden ge mer detaljerade anvisningar på vilka generella metadatastandarder som kan användas för digitala leveranser till TAM-Arkiv.

6.2.29 Juridisk information (LegalXML...)

Det finns lite olika försök att skapa XML-format för lagring och utbyte av juridisk information. Bland det mest spridda är LegalXML. LegalXML är en uppsättning av XML-format för att hantera juridiska dokument såsom stämningsansökningar, domar, juridiska citat, avskrifter, kontrakt med mera. LegalXML är skapat utifrån juridiska förhållanden i USA, vilket gör att vokabuläret antagligen måste anpassas kraftigt för att användas i Sverige.

I Sverige har funnits ett antal olika arbeten och projekt som studerat olika aspekter i rättskedjan, exempelvis Corpus Legis på 1990-talet, LISA Network (Legal Information Standards Action Network), Regeringskansliets rättsinformationsprojekt och Rättsväsendets Informationsförsörjning (RIF). Inget av dessa projekt har dock än så länge utmynnat i något standardiserat och allmänt använt format för långtidslagring av juridisk information i Sverige.

Rekommendation: Det är mycket tveksamt om LegalXML är tillräckligt utvecklat eller anpassat till europeiska förhållanden för att fungera för långtidslagring i dagsläget. I dagsläget torde PDF/A-1 rekommenderas i första hand för juridisk information, i väntan på en svensk standard för området.

6.2.30 Systemutveckling

Inom systemutveckling finns det tusentals olika programmeringsspråk, dialekter och olika tekniska plattformar. Inga av dessa är idag stabila på lång sikt. ISO 19503 XML Metadata Interchange (XMI) är en standard för utbyte av UML-modeller.

Rekommendation: I dagsläget finns det inget programmeringsspråk som är tillräckligt stabilt för att fungera för långtidslagring. Java har tyvärr blivit för komplext i senare versioner för att fungera för långtidslagring. Leverantörsoberoende och öppna språk har dock en tendens till att vara mer stabila än format som är bundet till en leverantörs plattform. XMI kan möjligen användas för lagring av UML-modeller i upp till 10 år. Men lagring i XMI bör kompletteras med parallell lagring i PDF/A-1, eftersom XMI tillåter en viss frihet i tolkningen. TAM-Arkiv planerar att i framtiden studera vidare detta område tillsammans med andra organisationer.

6.2.31 Artiklar och böcker

International Press Telecommunications Council (IPTC) har tagit fram ett format för utbyte av artiklar med bilder kallat NewsML. Dessutom har IPTC tagit fram ett format för uppmärkning av själva texten kallat News Industry Text Format (NIFT). En text uppmärkt med NIFT ger större sökmöjligheter än en vanlig löptext. OAIS DocBook är ett XML/SGML-format för uppmärkning och lagring av framförallt tekniska manualer, men kan även användas för böcker och artiklar i allmänhet. OAIS har tagit fram en förenkling av DocBook kallat OAIS Simplified DocBook. Text Encoding Initiative (TEI) används för att märka upp, kommunicera och lagra texter inom framförallt humaniora, lingvistik och socialvetenskaper. Även TEI kan användas för texter generellt. National Library of Medicine i USA har tagit fram standarden NLM Journal Archiving and Interchange för uppmärkning, lagring och utbyte av artiklar med mera till facktidskrifter. Ett annat format som förekommer inom förlagsindustrin för att hantera e-böcker är EPUB (Electronic PUblication), vilket används för att märka upp, kommunicera och lagra skönlitterära verk och viss facklitteratur. EPUB förvaltas av International Digital Publishing Forum (IDPF) och har ofta filändelsen .epub. EPUB ersätter den äldre standarden Open eBook/OEB/OEBPS.

Rekommendation: NIFT, NewsML, TEI och DocBook/Simplified DocBook bedöms vara tillräckligt stabila och utbredda för att fungera för långtidslagring. NLM Journal Archiving and Interchange och EPUB fungerar sannolikt även de för långtidslagring. Om man vill vara säker på att även bevara ursprunglig layout ska man använda PDF/A-1, exempelvis genom parallell lagring.

6.2.32 Personalrelaterad information

HR-XML är en uppsättning XML-vokabulärer för kommunikation av personalrelaterad information såsom CV, lönelistor, tidrapporter, anställningskriterier, personnamn, postadresser et cetera. HR-XML är i första hand utformat för förhållanden i USA.

Rekommendation: HR-XML-vokabulären är relativt omfattande och är inte specifikt anpassat till europeiska förhållanden. Dessutom är det huvudsakligen till för kommunikation och inte för lagring. Delar av HR-XML verkar dock vara tillräckligt stabilt för att kunna lagra data i åtminstone 5-10 år. Alternativet är PDF/A-1. Delar av HR-XML, som exempelvis adresser och personnamn, kan användas i kombination med andra standarder.

6.2.33 Digitala signaturer

I nuläget går det inte att garantera giltigheten för digitala signaturer över längre tid, eftersom en digital signatur egentligen bara är en kryptonyckel med en begränsad livslängd. En typisk livslängd för en digital signatur med normal nyckellängd är 5-10 år. Därefter är den inte att beteckna som giltig, eftersom den går att dekryptera relativt lätt. För att garantera läsbarhet och kontrollerbarhet för en digital signatur skulle det även krävas att kontexten till en digital signatur bevaras i form av information om CA (Certificate Authority), spärllistor CRL (Certificate Revocation List) och övrig PKI-struktur. Detta är något som är mycket svårt och dyrbart att göra i praktiken. På längre sikt är det möjligt att sveriges regering kommer att införa en nationell notariatstjänst som regelbundet kan omsignera och intyga att handlingar inte har förfalskats. Ett alternativ till en nationell digital notariatstjänst vore en gemensam inom EU. En notariatstjänst skulle lösa många av de nuvarande problemen med långtidslagring av digitalt signerade handlingar.

W3C XML Signature eller IETF RFC2315 PKCS #7 är de standarder som brukar användas för att signera och lagra digital information. Vid användande av PKCS#7 så finns den signerande partens "public key" antingen 1) med i ett certifikat eller 2) så finns bara en referens till specifik "public key" bestående av utfärdarens namn och serienummer för nyckeln. Denna information finns under "SignerInfo/issuerAndSerialNumber".¹³

¹³ Den information som finns under SignerInfo är version, issuerAndSerialNumber, digestAlgorithm, authenticatedAttributes, digestEncryptionAlgorithm, encryptedDigest och unauthenticatedAttributes.

Märk att det finns många lagar som kräver ”egenhändigt undertecknande” och som därför inte godkänner digitala signaturer. Exempel på lagar där det förekommer krav på egenhändigt undertecknande är Aktiebolagslagen, Jordabalken, Ärvdabalken, Rättegångsbalken, Konkurslag och Förvaltningslagen.

Ibland kallas digitala signaturer även för elektroniska signaturer. Oftast används dessa ord som synonymer. Rent tekniskt torde termen digital signatur vara mer korrekt. Ska man vara ännu mer exakt kan man säga validering genom asymmetrisk kryptering/dekryptering med kontrollsummeberäkning, vilket knappast någon orkar säga i praktiken.

Rekommendation: Information som ska långtidslagras bör lagras utan digital signatur och utan kryptering. Ett vanligt sätt att långtidslagra en signerad handling är att avsignera den och logga metadata om vad som avsignerats, inklusive information om den ursprungliga signaturen. För att minska risken för förvanskning är det extra viktigt att i detta fall ha arbets- och informationsprocesser som är informationskvalitetssäkrade.

Om man har ett mycket starkt behov av att kunna binda en informationsmängd till en specifik person/organisation, eller kunna bevisa med kryptonyckel att informationsmängden inte har manipulerats maximalt 5 - 10 år efter signering, kan man lagra samma informationsmängd parallellt både 1) med och 2) utan digital signatur och kryptering. Det vill säga att man parallelllagrar informationen, där 1) den signerade informationen används för juridiskt syfte i det kortare perspektivet och 2) den dekrypterade informationen med bortagen signatur används för att säkra den långsiktiga läsbarheten.

Notera att det finns andra metoder än digitala signaturer för att garantera äkthet hos information, eller för att binda en informationsmängd till en specifik person/organisation vid ett specifikt tillfälle och aktivitet, exempelvis att Anders Andersson godkände en utbetalning à 20 MSEK till ett konto i Bahamas år 2010 som råkade tillhöra hans hustru Anna Andersson. Andra metoder som man kan använda är omsignering, centraliserad timestamping, central notariatstjänst, parallell lagring av kopior (”rymdskyttele redundans”) och kvalitetssäkrade rutiner. Alla dessa metoder har sina styrkor och svagheter och uppfyller olika kravbilder. Tyvärr är detaljerna allt för omfattande för att få plats i detta dokument. Eventuellt kommer TAM-Arkiv i ett längre perspektiv att ta fram en separat rekommendation för långtidslagring av dokument med större krav på garanterad äkthet och juridisk bindning till en person/organisation, såsom vid exempelvis kontraktsskrivning.

6.2.34 ”Wrappers”-format (ASF, ISO_BMFF, RIFF, MXF, WARC, ZIP, TAR...)

”Wrappers” fungerar som en form av ryska dockor som packar in andra format i ett eller flera skikt. RIFF (Microsoft Resource Interchange File Format) och ASF (Microsoft Advanced Systems Format) kan exempelvis packa in olika multimediaformat. RIFF togs fram av IBM och Microsoft i början av 1990-talet. RIFF har ingen specifik filändelse utan filändelsen beror på vilket format som paketeras in såsom video (”.avi”), MIDI (”.rmi”), ljud (”.wav”) et cetera. AVI (Audio Video Interleaved) kan i sin tur paketera andra format såsom MPEG-, Xvid, DivX, JPEG- och WAVE-filer. ASF används huvudsakligen för att packa in WMA-filer

(Windows Media Audio) och WMV-filer (Windows Media Video). Microsoft äger formatet ASF, och typiska filändelser är .asf, .wma och .wmv. ISO Base Media File Format definieras i standarderna ISO 14496-12:2003 och ISO 15444-12:2004 och används för att paketera JPEG2000- och MPEG-4-filer. ISO Base Media File Format har beroenden till patent från flera olika företag. Apples ”wrapper”-format heter Quicktime och används för att lagra och ”streama” video, ljud, texter, grafik med mera. De vanligaste filändelserna för Quicktime är .qt och .mov. Som kuriositet kan nämnas att MPEG-4 har utvecklats utifrån en äldre version av Quicktime. MXF (Material Exchange Format) från Pro-MPEG Forum är ett wrapperformat för att packa in professionell video.

ISO 28500 WARC (Web ARChive) är ett format ursprungligen framtaget av Internet Archive för lagring av insamlade webbsidor. Formatet ZIP skapades ursprungligen år 1986 av Phil Katz för verktyget PKZIP. ZIP finns numera i ett antal olika versioner och anpassningar. ZIP-formaten används för att både komprimera och paketera. ZIP används bland annat i ODF- och JAR-filer. Ett annat vanligt paketeringsformat är TAR (Tape ARChive) som oftast återfinns inom UNIX/LINUX-världen. TAR används oftast i kombination med olika komprimeringsalgoritmer.

Rekommendation: TAM-Arkivs generella rekommendation är att man bör undvika ”wrappers” om det går. Detta eftersom många ”wrappers” har leverantörsberoenden och formaten dessutom kan vara ganska instabila, med undantag för ISO Base Media File Format, WARC, ZIP64 och liknande wrappers. Å andra sidan är en mycket stor del av filerna inom det audiovisuella området inpaketerade i wrappers, så användningen av wrappers kan vara svår att undvika i många praktiska sammanhang. Detta bland annat på grund av att många filformat bara har den ”inre” strukturen standardiserad, och inte den ”yttre” lagringsmässiga.

7. Vanliga frågor och svar (FAQ)

7.1 Allmänt

Vad är skillnaden mellan arkivläggning/arkivexport och back-up-tagning?

Arkivläggning/arkivexport innebär förenklat uttryckt att paketera om en viss informationsmängd till relativt beständiga format med tillhörande metadata. Metadata behövs för att kunna sköta framtida informationsförvaltning och återsökning, och även beskrivande kontextuellt metadata behöver medfölja för att förklara den ursprungliga informationsmängden. Arkivläggning brukar ske till plattformsoberoende format i så hög grad som möjligt.

Back-up-tagning å andra sidan brukar göras till plattformsoberoende format och används endast för att säkra information i databaser/filsystem/fysiskt media i allt från några dagar till några år. En back-up-kopia brukar oftast endast kunna läsas inom några år från när den skapades och den brukar vara hårt bunden till ursprunglig systemmiljö. En back-up-kopia fungerar därför inte för långtidslagring.

Vad är skillnaden mellan bevarande och gallring?

Termen ”bevarande”¹⁴ används i den svenska arkivteorin med betydelsen att en viss information ska göras tillgänglig för all framtid. Bevarande ställs i motsats till termen ”gallring”. Gallring innebär att en viss informationsmängd ska förstöras efter en viss tid och efter vissa specifika regler. Gallring kan även innefatta borttagande av 1) ursprungligt beteende såsom sökmöjligheter och sammanställningsmöjligheter, 2) audiovisuell presentation såsom ursprunglig bildmässig och eventuell ljudmässig presentation, 3) relationer till ursprunglig kontext eller 4) ursprungligt metadata.¹⁵ Syftet med gallring är oftast att få bättre prestanda i ett system, underlätta återsökning genom att den totala mängden information minskar och uppfylla lagkrav. I de flesta praktiska sammanhang brukar bevarande bara innebära att information inte ska gallras. Se även rekommendation ”TAM 4:2010 REKOMMENDATION - Arkivfunktionalitetskrav för IT-system” för övergripande information runt systemfunktionalitet för gallring.

Vad är gallringsfrist?

”Gallringsfrist” är den tid som information ska finnas kvar innan den ska förstöras. Vanliga krav som påverkar gallringsfristen för en godtycklig informationstyp är krav från den egna

¹⁴ Termen motsvaras närmast av den engelska termen ”preservation”.

¹⁵ Ibland används de juridiska termerna ”handling” och ”potentiell handling”, där ”handling” avser alla ursprungliga fixerade handlingar som framställdes, medan ”potentiell handling” avser de handlingar som var möjliga att framställa i det ursprungliga systemet. Detta ämne är mer komplext än vad som kanske verkar och skulle lätt kunna fylla en doktorsavhandlingen på tre hundra sidor.

organisationen, lagkrav, krav från medlemmar/kunder, leverantörer, massmedia, allmänhet och framtida forskning.

Vad är digital arkivering och långtidslagring?

Termerna ”digital arkivering”¹⁶ och ”långtidslagring”¹⁷ brukar tyvärr inte förekomma i svenska lagar eller föreskrifter, i alla fall inte än så länge. Digital arkivering innefattar all form av arkivhantering av digitalt lagrad information. Långtidslagring avser själva lagringen av digital information (data med beskrivande metadata) inklusive framtida konverteringar och annan kvalitetsbevarande vård. Information som långtidslagras kan gallras i framtiden, vilket gör att långtidslagring inte innebär samma sak som bevarande.

Exempel 1 på långtidslagring: Ett projektunderlag lagras och vårdas i ett digitalt arkiv i tjuugo år innan underlaget gallras.

Exempel 2 på långtidslagring: En verksamhetsberättelse ska bevaras, vilket teoretiskt innebär att den vårdas och lagras tills en större meteorit slår ner på jorden eller motsvarande.

Vad är informationsförvaltning?

Informationsförvaltning (Information Management) är konsten att hantera information inom en organisation på ett effektivt sätt och under kontrollerade former, så att korrekt och nödvändig information alltid finns tillgänglig. Informationsförvaltning inbegriper områden såsom scanning, digital arkivering, innehållshantering, information life cycle management, gallring, strukturering av information, metadata, informationsmodellering med flera. För den som är orienterad i the Zachman Framework kan nämnas att informationsförvaltning rör dimensionen ”Data – Vad?”. Se även TAM-Arkivs rekommendation ”TAM 2:2008 REKOMMENDATION – Informationsförvaltning”.

16 Den engelska termen ”digital archiving” innefattar arkivhantering av både material som ska 1) bevaras eller 2) gallras först på längre sikt. Andra engelska termer som har mer betoning bevarandeaspekten är ”long-time preservation” och ”digital preservation”.

17 Den engelska termen ”long-time storage” kommer från IT-världen (”storage”-sektorn) och avser specifikt lagring av data med tillhörande metadata på tiotals år sikt, till skillnad från ”vanlig storage” som brukar röra sig om maximalt tio år.

7.2 Frågor runt format för långtidslagring

Vilken upplösning ska man använda vid scanning av dokument?

Scanning av dokument bör ske i minst 300 dpi. Fotografier kan behöva scannas i minst 400 dpi gråskala/färg för att kunna få ett bra underlag till tryck. För högkvalitativa konsttryck bör 600 dpi användas.

Ska man zippa filer som ska arkiveras för att spara plats?

Nej, all form av komprimering, inkl. ”zippning”, av filer bör helst undvikas vid långtidslagring. Komprimering ökar risken att det inte går att läsa filerna i framtiden. Notera att i vissa fall ingår dock paketering med någon form av ZIP-container i själva filformatet.

Ska man kryptera filer som ska långtidslagras om man vill begränsa tillgängligheten?

Nej, filer som ska långtidslagras bör inte krypteras, eftersom detta kraftigt ökar risken att de inte går att läsa eller vårda i framtiden. Tillgängligheten till filer bör istället regleras genom behörighetskontrollsystem och kvalitetssäkrade manuella rutiner. Om man, vid långtidslagring, bedömer att information måste krypteras på grund av hög risk för skada/men, bör man istället överväga att gallra den. Se även avsnitt om digitala signaturer.

Varför kan man inte använda leverantörsformat för långtidslagring?

Det finns en stor uppsjö av olika leverantörsformat såsom formaten runt Quark Express, WordPerfect, PageMaker, Microsoft Word, Microsoft Excel, Microsoft PowerPoint, Microsoft Access, Adobe Photoshop, Mapinfo, Filemaker, Macromedia Flash, Real Media, AutoCAD, Adobe Acrobat med flera. Notera att även Adobe PDF (inte att förväxla med ISO PDF/A-1 eller ISO PDF/E) är ett leverantörsformat. Dessa format brukar dessutom finnas i olika versioner och profiler som förändras över tiden. Av fler skäl ska man inte använda denna typ av format för långtidslagring, bland annat för att leverantörsformat förändras (nya ”uppdaterade” versioner kommer ständigt), är ofta inte bakåtkompatibla, har ogenomtänkta strukturer som vuxit fram utifrån tillfälliga behov, kan innehålla kod och dessutom i många fall skyddas av patent. Detta gör att leverantörsformat oftast har en kort livslängd, under 5-10 år. Leverantörsformat går dock utmärkt att använda i det korta perspektivet, säg under 5 år. Notera att ett format kan vara godkänt av oberoende standardorganisation och ändå vara att betrakta som ett leverantörsformat om det är bundet till en specifik leverantörs produkt.

Vad är skillnaden mellan komprimering som är lossless respektive lossy?

Komprimering används för att minska storleken på filer genom användande av olika komprimeringsalgoritmer. Generellt bör man vid långtidslagring använda komprimeringsalgoritmer som inte förstör information, så kallad lossless-komprimering. Om man använder destruktiv komprimering (lossy) bör man ha en låg grad av destruktion.

Vad är skillnaden mellan rasterad och vektoriserad grafik?

Rasterad grafik bygger upp bilder med punkter, pixlar. Vektoriserad grafik använder geometriska (och i vissa fall fraktala) kurver och ytor för bygga upp bilder. Vektoriserad grafik används framförallt inom CAD, GIS, simuleringar, 3D-spel och animeringar.

Vilka färgrymder ska man använda för långtidslagring?

Att en färg alltid ser likadan ut på olika skärmar, skrivare och efter scanning tillhör undantagen. Färghantering blir lätt komplex om man går in i detaljer. Bland annat därför att upplevelsen av färg beror på hur det mänskliga ögat är konstruerat i tappar och stavar, vilka har ett icke-linjärt beteende. Generellt bör man använda en färgrymd/ett färgschema som inte är beroende av endast en leverantör eller ett operativsystem för långtidslagring. Se exempelvis specifikationerna ICC.1:2004-10 och ICC.1:1998-09 från International Color Consortium, sYCC (IEC 61966-2-1) och sRGB.

Vad innebär parallell lagring?

I vissa sammanhang såsom exempelvis inom systemutveckling och CAD kan den funktionalitet som erbjuds av långtidslagringsformat inte vara tillräcklig för den egna verksamhetens behov. I dessa fall kan man behöva lagra en arbetskopior i originalformat parallellt med en kopia i långtidslagringsbeständigt format. Arbetskopian är oftast i ett leverantörsspecifikt format, har en kortare livslängd, men har funktionalitet som inte finns i långtidslagringsformat. Kopior i långtidslagringsformat används å andra sidan för att garantera att i alla fall huvuddragen i informationen finns kvar.

8. Revisionshistorik

<i>Revision</i>	<i>Datum</i>	<i>Författare</i>	<i>Kommentar</i>
'1.0'	'2010-03-11'	Magnus Wåhlberg.	Första revisionen av rekommendationen för data-, teckentabell- och filformat.

Påpekanden om eventuella sakfel eller förslag på förbättringar av denna rekommendation kan skickas till magnus.wahlberg@tam-arkiv.se.

9. Bilagor

9.1 Bilaga 1, Ett urval av källor på World Wide Web

Specifika format:

Apache Torque (<http://db.apache.org/torque/>)

CDF (<http://cdf.gsfc.nasa.gov/>)

DocBook (<http://www.oasis-open.org/docbook/>)

EAC-CPF (<http://eac.staatsbibliothek-berlin.de/>)

EAD, METS, MIX, PREMIS, textMD (<http://www.loc.gov/standards/>)

ebXML (<http://www.ebxml.org/>)

ECMA-standarder (<http://www.ecma-international.org/>)

EPUB (<http://www.openebook.org/>)

GML (<http://www.opengeospatial.org/>)

HDF (<http://www.hdfgroup.org/>)

HR-XML (<http://www.hr-xml.org/>)

ISO-format (<http://www.iso.org/>)

JPEG och JPEG 2000 (<http://www.jpeg.org/>)

LegalXML (<http://www.legalxml.org/>)

NewsML, NITF (<http://www.iptc.org/>)

NLM Journal Archiving and Interchange (<http://dtd.nlm.nih.gov/>)

OAGIS (<http://www.oagi.org/>)

OASIS-format (<http://www.oasis-open.org/>)

PNG (<http://www.libpng.org/>)

PRISM (<http://www.prismstandard.org/>)

SIE-XML (<http://www.sie.se/format/siexml.asp>)

SveFaktura (<http://www.svefaktura.se/>)

U3D (<http://www.ecma-international.org/>)

W3C Date and Time Format (<http://www.w3.org/TR/NOTE-datetime>)

W3C Working with Time Zones (<http://www.w3.org/TR/2005/NOTE-timezone-20051013/>)

W3C-format (<http://www.w3.org/>)

WebCGM (<http://www.cgmopen.org/>)

VERS (<http://www.prov.vic.gov.au/vers/>)

WfMC-format (<http://www.wfmc.org/>)

X3D (<http://www.web3d.org/x3d/>)

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

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

DPE, DigitalPreservationEurope (<http://www.digitalpreservationeurope.eu/>)

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

PROTAGE (<http://www.protage.eu/>)

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

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

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

<http://www.ltu.se/ies/ldb-centrum/>



Titel: Rekommendation – Format för långtidslagring

Datum: 2010-03-11

Version 1.0

Översikt över vanliga filformat från Library of Congress:

(<http://www.digitalpreservation.gov/formats/>)

The National Archives UK PRONOM:

(<http://www.nationalarchives.gov.uk/PRONOM/>)

Översikt över vanliga XML-vokabulärer vid OAIS Cover Pages:

(<http://xml.coverpages.org/>)

Översikt över vanliga XML-vokabulärer från Barry & Associates, Inc:

(http://www.service-architecture.com/xml/articles/xml_vocabularies.html)

9.2 Bilaga 2, Ett kortare urval av rapporter, utredningar och föreskrifter

David Hay, *Data Model Patterns – A Metadata Map*, (Oxford 2006).

Margaret Hedstrom red., *It's about time – research challenges in digital archiving and long-term preservation*, (Washington 2003).

Stephan Heuscher, Stephan Järman, Peter Keller-Marxer och Frank Möhle, “Providing Authentic Long-term Archival Access to Complex Relational Data”, (European Space Agency Symposium “Ensuring Long-Term Preservation and Adding Value to Scientific and Technical Data”, 2004, Frascati, Italy.)

Christer Johansson, *Digital arkivering vid Göteborgs universitet - Utredning rörande förekomst av allmänna handlingar i digitala format samt förslag till modell för bevarande*, (Göteborgs universitet 2006).

ISO, *Information technology - Database languages - SQL - Part 14: XML-Related Specifications (SQL/XML)*, (ISO/IEC JTC 1/SC32/WG3).

Susan Lazinger, *Digital Preservation and metadata*, (Englewood 2001).

Lena Lindbäck, *Bevarande av webbsidor – ett gemensamt projekt mellan LTU och LDB-centrum*, (LDB-centrum 2009).

Lena Lindbäck, *CODA-WEBB – SLUTRAPPORT*, (LDB-centrum 2009).

Göran Lindqvist, *CODA-META - Curation of Digital Assets – Metadata – SLUTRAPPORT*, (LDB-centrum 2008).

Paradigm: Workbook on Personal Digital Archives, (Bodleian Library, Oxford 2007).

RA-FS 2009:1 - Riksarkivets föreskrifter och allmänna råd om elektroniska handlingar.

RA-FS 2009:2 - Riksarkivets föreskrifter och allmänna råd om tekniska krav för elektroniska handlingar.

Phil Shaw, *A JDBC-oriented XML DTD for SQL result sets*, (ISO/IEC JTC 1/32: Data Management and Interchange/WG3: Database Languages, 2000).

SLBA-utredningen, Utbildningsdepartementet, *SOU 2004:53, Översyn av verksamheten och arbetsformerna vid statens ljud- och bildarkiv, Bilaga 1-4*.

9.3 Bilaga 3, 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