

**Skýrsla um stöðu þekkingar og færni á
langtímavarðveislu stafræns efnis**

**Björn Þór Jónsson
Margrét Eva Árnadóttir**

***Menntamálaráðuneyti
2007***

ÚTDRÁTTUR.....	3
1 INNGANGUR.....	4
1.1 VARÐVEISLA STAFRÆNS EFNIS	4
1.2 SKÝRSLAN	5
1.2.1 UMFANG.....	5
1.2.2 UPPBYGGING	5
1.2.3 TENGILÍÐIR.....	5
2 ÞRÓUN Á HLUTVERKI SKJALASAFNA OG BÓKASAFNA	6
2.1 MUNUR VEGNA FJÖLDA EINTAKA	7
2.2 MUNUR VEGNA EÐLIS OG UPPRUNA UPPLÝSINGA.....	7
2.2.1 OPIÐ EFNI	8
2.2.2 LOKAÐ EFNI.....	8
2.2.3 VEFEFNI.....	8
2.2.4 GAGNAGRUNNAR	9
2.3 SAMVINNA VARÐVEISLUSTOFNANA	9
2.3.1 SAMÞÆTTING BÓKA- OG SKJALASAFNA	10
2.3.2 SAMÞÆTTING VINNSLUKERFA OG VARÐVEISLUKERFA	10
3 YFIRLIT YFIR VARÐVEISLUFERLI.....	11
3.1 OAIS.....	11
3.2 VIRK NOTKUN EFNIS	12
3.3 AFHENDING/INNTAKA STAFRÆNS EFNIS	12
3.4 ÖRUGG GEYMSLA OG VARÐVEISLA STAFRÆNS EFNIS.....	13
3.5 AÐGANGUR OG BIRTING STAFRÆNS EFNIS.....	16
4 FORDÆMI.....	18
4.1 ÍSLAND	18
4.1.1 LAGAUMHVERFI	19
4.1.2 ÞJÓÐSKJALASAFN ÍSLANDS	21
4.1.3 LANDSBÓKASAFN ÍSLANDS - HÁSKÓLABÓKASAFN	22
4.2 ÁSTRALÍA	23
4.2.1 NATIONAL ARCHIVES OF AUSTRALIA (NAA).....	23
4.2.2 NATIONAL LIBRARY OF AUSTRALIA (NLA)	26
4.3 BRETLAND	27
4.3.1 THE NATIONAL ARCHIVES (NA).....	27
4.4 BANDARÍKIN.....	30
4.4.1 NATIONAL ARCHIVES AND RECORDS ADMINISTRATION (NARA)	30
4.5 DANMÖRK.....	30
4.5.1 STATENS ARKIVER	31
4.6 SVÍPJÓÐ	32
4.6.1 RÍKSARKIVET	32
4.6.2 LDB-VERKEFNIÐ.....	32
4.7 HOLLAND.....	33

4.7.1	KONINKLIJKE BIBLIOTHEEK (KB).....	33
4.7.2	NATIONAAL ARCHIEF.....	35
4.8	KANADA.....	35
4.8.1	LIBRARY AND ARCHIVES CANADA	35
4.9	EVROPUSAMBANDIÐ	35
4.10	ÖNNUR KERFI.....	36
<u>5</u>	<u>STAÐLAR.....</u>	<u>38</u>
5.1	FERLI	38
5.2	LÝSIGÖGN.....	38
5.3	SKJALASNIÐ.....	39
5.3.1	VAL Á SKJALASNIÐUM	39
5.3.2	PDF/A	40
<u>6</u>	<u>SAMANTEKT.....</u>	<u>41</u>
6.1	FJÁRMAGN.....	41
6.2	LAGALEG ATRIÐI	41
6.3	TÆKNILEG ATRIÐI.....	42
6.4	NIÐURSTAÐA	43
<u>7</u>	<u>HEIMILDASKRÁ</u>	<u>44</u>

Útdráttur

Langtímavarðveisla stafræns efnis er stórt verkefni sem mörg lönd huga nú alvarlega að, þar á meðal Ísland. Langtímavarsla safnefnis hefur ávallt verið umfangsmikið vandamál, en tæknilegt eðli stafræns safnefnis gerir langtímavörslu að sumu leyti enn erfiðari, auk þess sem magn stafræns efnis eykst gífurlega hratt um þessar mundir. Á móti kemur þó að tæknilegur munur á safnefni þjóðbókasafna og þjóðskjalasafna er að minnka og segja má að báðar safnategundirnar standi frammi fyrir svipuðum vanda. Þess vegna má leiða rökum að því að aukin samvinna slíkra safna geti eflt starfsemi þeirra.

Langtímavarðveisla stafræns efnis hefur verið rannsökuð í nokkurn tíma. Ýmislegt hefur áunnist og mikil undirbúningsvinna er að baki. OAIS staðallinn hefur náð fótfestu sem staðall fyrir varðveisluferli, ýmsir staðlar hafa verið þróaðir fyrir skjalasnið (t.d. PDF/A) og lýsigögn (t.d. PREMIS) og mörg kerfi hafa verið þróuð með langtímavörslu að markmiði. Þrátt fyrir það er mikið verk óunnið og langt í land með að vandamálið teljist leyst.

Skýrsla þessi er unnin að beiðni vinnuhóps á vegum menntamálaráðuneytis um tæknileg atriði varðandi vörslu rafrænna gagna. Hlutverk skýrsluhöfunda var taka saman og meta núverandi stöðu á langtímavarðveislu stafræns efnis í heiminum og finna mögulegar fyrirmyndir fyrir Ísland. Megináherslan var lögð á að finna góðar fyrirmyndir um tæknilegar lausnir við varðveislu stafræns efnis á þeim bóka- og skjalasöfnum sem eru í fremstu röð í þessum málum í heiminum. Auk umfjöllunar um mögulegar fyrirmyndir er einnig fjallað að nokkru marki um ferli við varðveislu, gagnasnið, lýsigögn og fleiri mikilvæga þætti langtímavarðveislu.

Þegar litið er til þeirra landa sem fjallað er um í þessari skýrslu, er það álit skýrsluhöfunda að The National Archives í Bretlandi og Koninklijke Bibliotheek í Hollandi séu bestu fordæmin í nágrannalöndum okkar. The National Archives hefur skilgreint mjög heilsteypta, sérsmíðaða högun (e. architecture) fyrir langtímavarðveislu. Einnig er tenging kerfisins við lokað net stjórnsýslunnar þar í landi mjög áhugaverð. Koninklijke Bibliotheek hefur einnig komið upp áhugaverðu kerfi. Athyglisverðust er þó vinna þeirra við eftirfylgni með skjalasniðum. Norðurlandþjóðir vinna einnig að þessum málum og það verður einkum áhugavert að fylgjast með framgangi Svía á næstu árum, þegar LDB-verkefni þeirra um menntun og rannsóknir á sviði langtímavarðveislu fer að skila árangri. Loks verður áhugavert fyrir Ísland að reyna að tengjast evrópsku rannsóknarverkefnunum Planets og CASPAR

1 Inngangur

Skýrsla þessi er unnin að beiðni vinnuhóps á vegum menntamálaráðuneytis um tæknileg atriði varðandi varðveislu rafrænna gagna. Í hópnum voru Halla Björg Baldursdóttir forsætisráðuneyti, formaður, Bjarni Þórðarson, Þjóðskjalasafni Íslands, Guðmundur Kjærnested, menntamálaráðuneyti, Ingibjörg Sverrisdóttir, Landsbókasafni Íslands - Háskólabókasafni og Kjartan Ólafsson, Fakta.

1.1 Varðveisla stafræns efnis

Bóka- og skjalasöfn hafa sinnt því hlutverki að varðveita útgefið efni og skjöl og þar með varðveitt menningararf landsins. Að mörgu þarf að huga þegar efni er varðveitt til langs tíma og hafa umhverfisþættir hingað til verið aðaláhyggjuefnið: Raki, hitastig, eldur, vatn ofl. Auk þessa hefur fjöldaframleiðslan gert erfitt fyrir að varðveita efni, kúlupennablek eyðist mjög hratt – jafnvel á 10 árum, pappír hefur hærra sýrustig en æskilegt er og eyðir því út frá sér, plastmöppur, hefti og annað þess háttar skemmir líka efnið.

En nú hefur nútímatækni haft í för með sér fleiri áhyggjuefni þegar kemur að varðveislu. Stafrænt efni er töluvert öðruvísi en prentað efni og eðli þess gerir það að verkum að það getur verið töluvert snúið að varðveita það. Vert að velta fyrir sér muninum á stafrænu efni og prentuðu efni og bendir Steenbakkars (2005)¹ aðallega á þrjú atriði í því sambandi:

- Prentað efni er raunlægt og áþreifanlegt (e. physical object). Stafrænt efni er hins vegar röklegt og óáþreifanlegt (e. logical object) sem geymt er á raunlægum miðli.
- Hið mennska auga getur samstundis lesið prentað efni en það þarf að notast við hugbúnað til að lesa stafrænt efni af hinum raunlæga miðli.
- Prentað efni getur varðveist fyrir slysi og án ásetnings, og haldist læsilegt í langan tíma án þess að sérstakra aðgerða sé þörf til að tryggja það. Frekar langsótt er að slíkt geti átt við stafrænt efni.

Langtímavarsla stafræns efnis án aðgengis hefur engan tilgang. Mikilvægt er að huga vel að aðgengi eigi að varðveita stafrænt efni til frambúðar. Tryggja þarf annars vegar að ávallt séu til eintök af efninu á aðgengilegum raunlægum miðli, og hins vegar að ávallt sé til hugbúnaður sem getur lesið efnið og birt á læsilegan hátt.

Það má telja nánast útilokað að hægt sé að finna eina endanlega lausn fyrir varðveislu stafræns efnis. Tækninni fleytir hratt áfram; bæði vél- og hugbúnaður úreldist mjög hratt og gerir það að verkum að erfitt er að varðveita stafrænt efni í upprunalegu formi og umhverfi. Aðferðafræði sú sem flestir nota í dag snýst því um tvo þætti:

- Að tryggja gæði efnis sem varðveita á, með efnissniðum og lýsigögnum, þannig að líkur séu á að það varðveitist vel.
- Að finna skammtímalausnir sem varðveita efnið á öruggan máta í 5-10 ár og aðferðir við uppfærslu kerfa, miðla og efnis til að koma því frá einni skammtímalausn yfir á aðra.

Öll þessi atriði og mörg fleiri hafa ýtt af stað mikilli rannsóknarvinnu um allan heim hjá bæði bókasöfnum og skjalasöfnum, þar sem leiðarljósið hefur verið hvernig hægt sé að tryggja langtímarvarðveislu stafræns efnis. Þessi rannsóknarvinna er mislangt á veg komin og er engan veginn lokið.

¹ Steenbakkars, J.F. *Digital archiving in the Twenty-First century*. 2005. Bls. 35

1.2 Skýrslan

1.2.1 Umfang

Hlutverk skýrsluhöfunda var taka saman og meta núverandi stöðu á langtímarvarðveislu stafræns efnis í heiminum og finna mögulegar fyrirmyndir fyrir Ísland.

Ljóst er að viðfangsefni skýrslunnar er því mjög viðamikill og reyndist nauðsynlegt að setja verkefninu ákveðnar skorður til að halda því innan tíma- og fjármagnsramma. Megináherslan var lögð á að finna góðar fyrirmyndir um tæknilegar lausnir við varðveislu efnis á þeim bóka- og skjalasöfnum sem eru í fremstu röð í þessum málum í heiminum. Fjöldmörg rannsóknarverkefni hafa verið unnin sem snúa að langtímarvarðveislu stafrænna gagna, en umfangsins vegna var ákveðið að fjalla ekki um þau í þessari skýrslu.

Auk umfjöllunar um mögulegar fyrirmyndir úti í heimi er einnig fjallað að nokkru marki um ferli við varðveislu, gagnasnið, lýsigögn og fleiri mikilvæga þætti langtímarvarðveislu.

Umfang skýrslunnar miðar við safnefni eftir að það er afhent varðveislustofnun og er því orðið óvirkt.

1.2.2 Uppbygging

Ákveðið var að hefja skýrsluna á að fjalla um þróun á hlutverki skjalasafna og bókasafna og fara yfir þann mun sem er á safnkosti þeirra og mögulega samvinnu. Þar næst tekur við umfjöllun um varðveisluferli stafræns efnis sem fellur undir umfang skýrslunnar. Næsti kafli lýsir stöðu verkefna í þeim löndum sem valin voru til frekari skoðunar. Þá tekur við umfjöllun um helstu staðla sem eiga við um ferli, lýsigögn og skjalasnið. Að lokum eru niðurstöður skýrslunnar dregnar saman. Allítarleg heimildaskrá fylgir, sem sýnir þær heimildir sem stuðst var við við gerð skýrslunnar.

1.2.3 Tengiliðir

Eftirfarandi aðilar tóku vel á móti fyrirspurnum okkar og tóku sér tíma til að svara öllum okkar spurningum og kunnum við þeim okkar bestu þakkir fyrir:

- Þjóðskjalasafn Íslands: Baldur Már Bragason, Eiríkur G. Guðmundsson og fleiri
- Landsbókasafn Íslands - Háskólabókasafn: Kristinn Sigurðsson, Þorsteinn Hallgrímsson og fleiri
- National Archives of Australia, Ástralíu: Marian Hoy
- National Library of Australia, Ástralíu: Paul Koerbin
- The National Archives, Bretlandi: Adrian Brown
- Statens Arkiver, Danmörku: Anders Bo Nielsen
- Riksarkivet, Svíþjóð: Jonas Palm
- LDB verkefnið, Svíþjóð: Jan Aspenfjäll
- Koninklijke Bibliotheek, Hollandi: Hilde van Wijngaarden
- Nationaal Archief, Hollandi: Jacqueline Slats

Líka var reynt að hafa samband við Kanada og Bandaríkin, en engin svör bárust:

- Library and Archives Canada, Kanada: René Paquet
- National Archives and Records Administration, Bandaríkjunum: ERA Program

Auk þess hefur vinnuhópur um tæknileg atriði varðandi varðveislu stafræns efnis verið okkur innan handar á allan hátt, og kunnum við þeim bestu þakkir fyrir.

2 Þróun á hlutverki skjalasafna og bókasafna

Hlutverk skjalasafna og bókasafna hefur m.a. mótast af því efni sem söfnin hafa varðveitt. Á Íslandi hefur hlutverk tveggja stærstu safnanna verið skilgreint af lögum um Þjóðskjalasafn Íslands nr. 66/1985 og um Landsbókasafn Íslands - Háskólabókasafn nr. 71/1994.

Samkvæmt lögum nr. 66/1985 er hlutverk ÞÍ m.a. að:

- Heimta inn og varðveita skjöl þeirra aðila sem eru afhendingarskyldir samkvæmt 5. gr. þessara laga,
- Líta eftir skjalasöfnum afhendingarskyldra aðila, láta þeim í té ráðgjöf, gefa út leiðbeiningar um skjalavörslu og tölvuskráningu og ákveða ónýtingu skjala sem ekki er talin ástæða til að varðveita til frambúðar,
- Gangast fyrir fræðslu um skjalavörslu fyrir fólk sem á að annast skjalavörslu í opinberum stofnunum, svo sem með námskeiðum og leiðbeiningarritum,
- Skrásetja varðveitt skjalasöfn og gefa út prentaðar eða fjölritaðar skrár um þau til leiðbeiningar um notkun þeirra,
- Koma upp handbókasafni um skjalfræðileg og sagnfræðileg efni,
- Líta eftir starfsemi héraðsskjalasafna og annarra skjalavörslustofnana sem varðveita opinber skjöl,
- Leiðbeina safngestum um notkun varðveittra skjala og heimilda safnsins, vísa safngestum á heimildir eftir því sem kostur er, efla þekkingu á þjóðarsögunni og stuðla að rannsóknum á henni,
- Halda opnum lestrarsal fyrir almenning þar sem unnt sé að sinna fræðistörfum og færa sér í nyt varðveitt skjöl og aðrar heimildir safnsins,
- Safna öðrum skráðum heimildum þjóðarsögunnar innanlands og utan, þ. á m. ljósritum og öðrum afritum ef frumheimilda er eigi kostur, og stuðla að varðveislu þeirra,
- Reka viðgerðar- og bókbandsstofu og taka í þjónustu safnsins þá tækni sem á hverjum tíma telst æskileg til þess að fullkomna starfsemi þess.

Hlutverk Lbs-Hbs er skilgreint í lögum nr. 71/1994, en stór hluti snýr að söfnun íslensks efnis. Hlutverk safnsins er m.a. eftirfarandi:

- Að viðá að sér gögnum í prentuðu formi eða á öðrum miðlum, skrá þau og búa í hendur notendum.
- Að þaulsafna íslenskum gögnum, m.a. með viðtöku skylduskila, svo og að afla erlendra gagna er varða íslensk málefni.
- Að varðveita handritasöfn þau sem stofnað hefur verið til, sbr. 2. mgr. 14. gr., vinna að frekari söfnun og rannsóknum íslenskra handrita og samsvarandi efnis á nýrri miðlum. Hið sama á við um hliðstætt erlent efni sem varðar Ísland.
- Að tryggja sem best viðhald og varðveislu safnkostsins. Í því skyni skal m.a. starfrækja bókbandsstofu, viðgerðarstofu og myndastofu. Þá skal taka frá eitt eintak af öllu efni sem berst í skylduskilum, undanskilja það allri venjulegri notkun og geyma tryggilega.
- Að starfrækja bókminjasafn.
- Að gera skrár um íslenskar bækur, handrit og hljóðrit, svo og eftir atvikum margvíslegar efnisskrár.
- Að halda uppi rannsóknum á sviði íslenskrar bókfræði og bóksögu og veita upplýsingar um íslenska bókaútgáfu.
- Að gefa safngestum kost á vinnuástöðu og sem greiðustum aðgangi að safngögnum.

- Að starfrækja samskrá bókasafna og láta bókasöfnum í té tölvu- og skráningarþjónustu eftir því sem stjórn bókasafnsins ákveður.
- Að stuðla að samræmingu starfshátta í íslenskum bókasöfnum, veita þeim faglega ráðgjöf og eiga við þau sem víðtækast samstarf.
- Að taka þátt í fjölþjóðlegu samstarfi á sviði rannsóknarbókasafna og upplýsingamála.

Af þessu má því sjá að þó bæði söfnin eigi að safna íslensku efni er eðli þess efnis mismunandi. Um er að ræða þrenns konar mismun. Í fyrsta lagi er efni safnanna flokkað á mismunandi hátt, en þar sem flokkun efnis hefur lítil áhrif á langtímarvarðveislu eru flokkunarkerfi ekki rædd frekar. Í öðru lagi er um að ræða mun sem orsakast af fjölda eintaka, sjá umfjöllun í kafla 2.1, og í þriðja lagi mun sem orsakast af eðli og uppruna upplýsinganna, sjá umfjöllun í kafla 2.2. Í kafla 2.3 er loks fjallað um mögulegt samstarf þessara stofnana.

2.1 Munur vegna fjölda eintaka

Almennt séð hafa bókasöfn varðveitt útgefin verk, en þetta er þó mismunandi milli landa. Á Íslandi gilda lög um skylduskil til safna nr. 20/2002 og segja þau til um þann fjölda eintaka efnis sem útgefendum ber að skila til safna til varðveislu. Um er að ræða birt efni, sem telst birt í skilningi laganna þegar eintök af því hafa verið löglega gefin út eða þegar það er með rétttri heimild gert aðgengilegt almenningi með öðrum hætti, svo sem ef það er flutt eða sýnt opinberlega eða er aðgengilegt almenningi á rafrænu formi um tölvunet. Mismunandi er hversu mörgum eintökum á að skila og fer það eftir efnismiðli, t.d. skal skila fjórum eintökum af útgefnum bókum en þremur af verkum gefnum út á rafrænu formi.

Tekið er fram í lögum um skylduskil til safna, nr. 20/2002, að varðveislusöfnin skulu leggja til hliðar eitt eintak og geyma sérstaklega, og hafa þau komið sér upp varaeintakasöfnum til að koma til móts við þetta. Þannig er talið að hægt sé að tryggja betur langtímarvarðveislu útgefna, íslenskra verka.

Eðli útgefna verka hefur verið þannig að þau eru útgefin í stóru upplagi og því er sjaldgæft að bókasöfn séu með síðasta þekkta eintakið af verkinu, þótt þess þekkist vissulega dæmi.

Skjalasöfn hafa aftur á móti almennt varðveitt efni sem aðeins er til í einu eintaki, eins og stjórnsýslugögn, einkabréf og fleira. Því hefur verið mjög mikilvægt að varðveita eintökin vel, þar sem þau eru óbætanleg.

Með tilkomu stafræns efnis má segja að safnkostur beggja safnategunda sé orðinn sams konar þegar litið er til miðilsins. Stafrænt efni er venjulega aðeins búið til í einu eintaki og oft er ekki einu sinni hægt að tala um eintök, eins og t.d. er tilfellið með vefsíður og vefrit. Hins vegar má auðveldlega búa til fleiri „eintök“ með því að afrita efnið og dreifing þess er einnig auðveldari. Þetta gerir það að verkum að mjög mikilvægt er að glata ekki síðasta eintakinu, líkt og á við um skjölin á skjalasöfnum, en á móti að auðvelt er að búa til fleiri eintök, sem er líkara því sem á við um bókasöfn.

2.2 Munur vegna eðlis og uppruna upplýsinga

Eðli upplýsinganna í safnefni bókasafna og skjalasafna er mismunandi og sá mismunur birtist aðallega þegar litið er á „opinleika“ efnisins. Flokka má efnið gróflega í „opið“ og „lokað“, og athuga skal að efni flokkast oft undir sitt hvorn flokkinn á mismunandi hluta lífsskeiðsins. T.d. er hægt að setja kvaðir á efni þannig að það sé lokað í ákveðinn tíma og opnað að þeim tíma loknum.

Vefefni og innihald gagnagrunna hafa talsvert aðra eiginleika en hefðbundin skjöl. Ákveðið var því að fjalla sérstaklega um þetta efni og er það gert í köflum 2.2.3 og 2.2.4.

2.2.1 Opið efni

Opið efni skilgreinum við sem það efni sem er aðgengilegt almenningi frá því það er afhent varðveislustofnun. Hvernig þeim aðgangi er háttáð getur verið mismunandi og geta lögin t.d. takmarkað aðgang að slíku efni. Höfundalögin nr. 73/1972 eru dæmi um það, en þau gera það að verkum að þó svo hægt sé að fá lánaðar bækur frá bókasafni, þá er safninu bannað að gera stafrænar útgáfur aðgengilegar á netinu. Einungis má veita aðgang að þeim í safninu sjálfu. Lögin gilda hvort sem um er að ræða formlega útgáfu eða óformlega útgáfu eins og tíðkast á netinu.

Opið efni sem vistað er á skjalasöfnum er efni sem er orðið 30 ára og á hvíla engar aðgangstakmarkanir. Stofnanir geta þó fengið undanþágu og afhent efni fyrr, sem gerir það að verkum að opið efni yngra en 30 ára getur verið aðgengilegt almenningi.

Nefna skal að söfnum getur borist efni frá einstaklingum, sem vilja gefa safninu persónuleg efnissöfn til varðveislu. Á þessum söfnum geta legið aukakvaðir hvað varðar aðgang hvernig svo sem eðli upplýsinganna sé. Lbs-Hbs varðveitir mörg söfn af þessum toga og þarf að taka tillit til óska gefenda safnanna. Þetta er dæmi um hvernig uppruni skjala getur haft áhrif á opinleika þeirra.

2.2.2 Lokað efni

Lokað efni er það efni sem ekki er aðgengilegt almenningi. Um getur verið að ræða skjöl og annað opinbert efni, og ýmislegt persónulegt efni. Þetta efni er mjög viðkvæmt og þarf því að stýra aðgengi að því vandlega. Hér kemur löggjöfin einnig við sögu og er hún umtalsverð. Upplýsingalögin nr. 50/1996 fjalla um upplýsingaskyldu opinberra aðila og einnig hvaða efni er undanþegið þeirri skyldu. Þessi lagaskilyrði geta orðið töluvert flókin og því þarf Úrskurðarnefnd um upplýsingamál oft að úrskurða í vafamálum er að þessu líta. Í stjórnarskrá Íslands eru einnig ákvæði varðandi aðgengi að því efni sem upplýsingalög ná ekki yfir. Lög um persónuvernd og meðferð persónuupplýsinga nr. 77/2000 koma líka hér við sögu, ásamt öðrum sértækum lögum eins og læknalögum. Þetta efni má því kalla lokað þar sem sérlega mikilvægt er að það komist ekki í hendur þeirra sem ekki mega sjá það.

Eins og nefnt var í upphafi kafla 2.2, þá er líka til efni sem er lokað í takmarkaðan tíma. Sá tími getur annars vegar ákvarðast af eðli og uppruna efnisins, og hins vegar sérákvæða sem hvíla á því.

2.2.3 Vefefni

Tilkoma Internetsins hefur haft veruleg áhrif á útgáfustarfsemi og upplýsingagjöf. Bæði skjalasöfn og bókasöfn þurfa að huga að vistun vefgagna, þar sem um er að ræða allt frá opnum almennum vefjum yfir í lokaða stjórnsýsluvefi.

Vefefni hefur ákveðnar sérþarfir hvað varðar langtíma varðveislu. Um er að ræða mikinn fjölda skjala og gríðarlegt gagnamagn, sem liggur ekki síst í myndum og öðru efni en texta. Vefir eru breytingum háðir og venjulega er ekki hægt að tala um endanlega útgáfu af efni, heldur þarf að fylgjast með breytingum. Þar að auki geta vefir breyst mjög hratt, þannig að ekki er raunhæft að vista allar útgáfur sums vefefnis, heldur verður að vista innihald vefjanna á ákveðnum fresti.

Birting vefgagna er einnig erfiðleikum háð, þar sem mjög margar tegundir skjalasniða eru í notkun, og mismunandi undirliggjandi tækni er notuð við útfærslu vefjanna. Vegna umfangs

Internetsins er ógjörningur að ætlast til umbreytingar skjala yfir á stöðluð snið. Því er sérstaklega erfitt að tryggja að vefefni sé birtanlegt þegar til lengri tíma litið.

Loks má þó nefna að oft er vefefni unnið upp úr gagnagrunnum stofnana, og í þeim tilvikum getur verið auðveldara að tryggja varðveislu upplýsinga með því að varðveita þá gagnagrunna, eins og rætt er hér að neðan.

2.2.4 Gagnagrunnar

Eins og sagði hér að ofan eru gagnagrunnar nýleg tegund efnis sem mögulegt er að varðveita í stafrænni langtímavarðveislu. Gagnagrunnar eru venjulega samsafn mismunandi taflna, sem hver um sig inniheldur margar smáar færslur. Færslur í mismunandi töflum geta tengst saman gegnum svokölluð vensl og þannig myndað eina heild sem dreifð er um gagnagrunninn. Í skattagrunni væri til dæmis ein lykilmærsla um einstakling í viðeigandi töflu, en aðrar töflur gætu innihaldið margar færslur um aðstandendur, laun og frádrætti, eignir, skuldir o.s.frv. Gagnagrunnar eru venjulega notaðir af forritum sem vinna efni upp úr mörgum töflum og birta, t.d. í skýrslum eða skjölum. Dæmi um slík skjöl er t.d. álagningarseðill frá skattstjóra.

Nokkur umræða hefur komið upp um varðveislu gagnagrunna umfram eða til viðbótar við annað efni, eins og skjöl þau sem unnin eru upp úr gagnagrunnunum. Að minnsta kosti þrjár ástæður liggja að baki því að vista gagnagrunna:

- Í fyrsta lagi rata ekki alltaf allar upplýsingar þeirra í skjöl, en geta samt verið gagnlegar við rannsóknir seinni tíma.
- Í öðru lagi getur úrvinnsla seinni tíma kallað á greiningu og samantekt gagna, og þá væri mjög tímafrekt að vinna upplýsingarnar úr skjölum.
- Í þriðja lagi eru skjölin miklu frekari á rými og þar sem um stöðluð skjöl er að ræða má spara rými með því að geyma frekar form skjalanna og grunnupplýsingarnar í gagnagrunnum.

Sú spurning getur þá vaknað hvort nægilegt sé að vista gagnagrunnana eingöngu, og sleppa því alfarið að vista skjölin. Svo er ekki, af nokkrum ástæðum: Í fyrsta lagi getur verið erfitt að tryggja að samspil gagnanna verði ávallt ljóst, þar sem oft er um að ræða samspil gagna, forrita og birtingarsniðs sem ræður útliti skjala. Í öðru lagi geta skjöl lýst ákvörðunum sem ekki eru skrásettar í gagnagrunnum. Í þriðja lagi getur óyggjandi form skjala skipt máli um sönnunargildi þeirra.

Nefna má að varsla skjala getur auðveldað afgangi á efni gagnagrunna. Þetta á einkum við þegar óskað er eftir einstökum skjölum úr safninu, eins og þegar einstaklingar koma og óska eftir afriti af sínum eigin skattskýrslum. Þannig gæti vel átt við að vista einstakar skattskýrslur hjá skjalasafni, þótt þær séu núna unnar beint upp úr gagnagrunninum hjá skattstjóra þegar um þær er beðið.

Í ljósi þessa er nauðsynlegt að skoða hvern gagnagrunn fyrir sig og ákveða hvaða töflur þarf að varðveita og hvaða skjöl.

2.3 Samvinna varðveislustofnana

Ljóst er að tæknilegur munur á safnaefni er að minnka með tilkomu stafræns efnis. Því má hugleiða mögulega samvinnufleti milli varðveislustofnana til að auðvelda skjalastjórn og langtímavarðveislu efnis. Einnig er vert að velta fyrir sér stöðu vinnsluferfa og varðveislukerfa.

2.3.1 Samþætting bóka- og skjalasafna

Hugmyndin um að sameina bókasöfn og skjalasöfn, þar sem báðar stofnanir sinni langtímarvarðveislu efnis á ýmsu formi, er ekki ný af nálinnu. Þessi leið hefur til að mynda nýlega verið farin í Kanada, þar sem National Library og National Archives sameinuðust árið 2004 í Library and Archives Canada.

Ekki er þó nauðsynlegt að sameina þessar stofnanir til að ná fram hagræðingu þegar kemur að langtímarvarðveislu stafræns efnis, heldur gæti samvinna leitt til hagræðingar án þess að breyta grundvallareðli stofnananna. Öll slík samvinna kallar á vandvirk vinnubrögð og góða undirbúningsvinnu, en ekki verður farið nánar út í þetta í þessari skýrslu.

2.3.2 Samþætting vinnslukerfa og varðveislukerfa

Vinnslukerfi eru kerfin sem eru í notkun í stofnunum þar sem unnið er með efnið á hverjum degi og þar sem efnið eru myndað. Þar á móti eru langtímarvarðveislukerfi þau kerfi sem notuð eru til að vista efnið til langs tíma og eru í umsjón varðveislustofnunar.

Sú spurning hefur vaknað hvort hægt sé að sameina vinnslukerfi og langtímarvarðveislukerfi í eitt kerfi. Ef þessar tegundir kerfa yrðu sameinaðar í eitt, myndu efnismyndarar strax geyma allt efni á einum miðlægum stað og leiðin úr vinnslu í varðveislu væri stutt. Hins vegar er eðli virks og óvirks efnis mjög mismunandi. Virkt efni er í stöðugri notkun, á meðan óvirkt efni er sjaldan eða aldrei notað. Gildir þá einu hvort um stafrænt eða annað efni er að ræða. Sér í lagi getur virkt efni verið breytingum og viðbótum háð, meðan óvirkt efni breytist ekki. Þar sem markmið langtímarvarðveislur er að viðhalda óbreyttu efni, er almennt ekki talið samrýmanlegt að hafa þetta tvennt saman í einu kerfi.

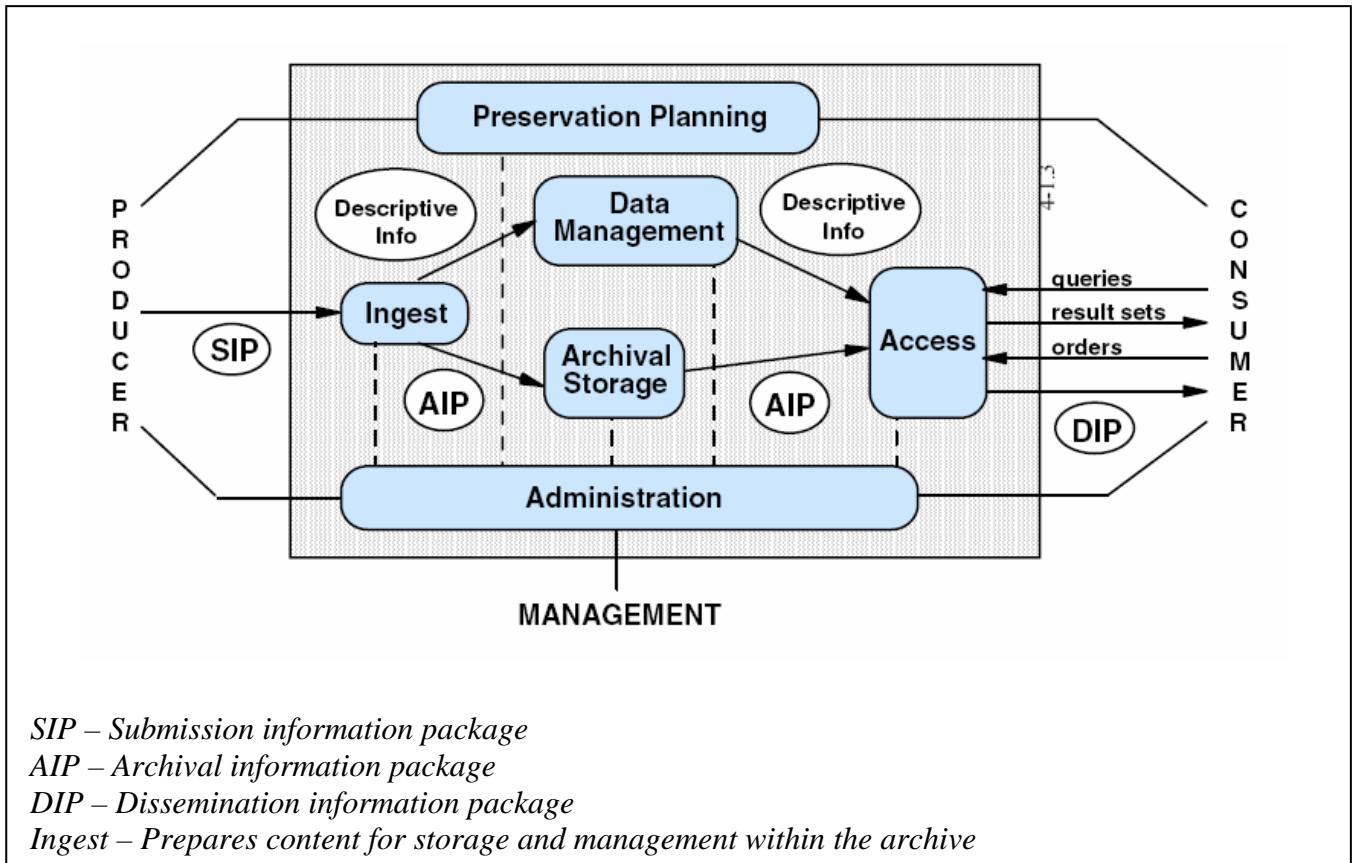
Hins vegar má mögulega taka leiðina sem valin er í Bretlandi til fyrirmyndar, en þar verður efni skilað mjög snemma til skjalavörslu hjá The National Archives. Safnið mun veita efnismyndurum aðgang að efninu yfir lokað net stjórnsýslunnar (GSI, Government Secure Intranet), en það er uppsett og rekið af fyrirtækinu Energis.² Þannig væri hægt að hlífa opinberum stofnunum við þeirri fyrirhöfn að varðveita efni til allt að 30 ára eins og nú er nauðsynlegt á Íslandi (með tilheyrandi lagabreytingu). Rétt er að benda á að þessi aðferð byggir fyrst og fremst á tilveru lokaðs nets stjórnsýslunnar (eða annarrar útfærslu á öruggri samskiptabraut milli stofnana), sem á að tryggja öruggan aðgang að innri kerfum og efni stjórnsýslunnar.

² eGovernment News. *UK: Improved version of the UK Government Secure Intranet launched.* 2004. <http://ec.europa.eu/idabc/en/document/2136>

3 Yfirlit yfir varðveisluferli

3.1 OAIS

OAIS (Open Archival Information System) staðallinn er ISO staðall sem snýr að varðveisluferli, þ.e. ferli stafræns efnis frá afhendingu til varðveislustofnunar til birtingar. Eftirfarandi mynd lýsir þeirri uppbyggingu sem nota þarf í slíkum kerfum³; í öllum löndum sem skoðuð voru er þessi staðall notaður í kerfum safnanna eða a.m.k. hafður til hliðsjónar við þróun þeirra. Með tilkomu staðalsins hefur komið ákveðið hugtakasafn sem gerir alla samvinnu auðveldari.



Mynd 3-1 : OAIS

Til að skýra varðveisluferlið er hentugt að líta á fjögur undirferli: Virk notkun efnis, afhending/inntaka efnis, örugg geymsla og varðveisla, og aðgangur og birting efnis. Þessi ferli verða skoðuð hér í þessum kafla, áður en einstök söfn eru skoðuð í kafla 4.

Þess má geta að Research Libraries Group (RLG) og National Archives and Records Administration (NARA) í Bandaríkjunum hafa gefið út drög að gátlista fyrir „áreiðanlegar stafrænar geymslur“ (e. trusted digital repositories).⁴ Þar sem enn er um drög að ræða hafa engin söfn farið formlega í gegnum slíka vottun, en drögin eru engu að síður áhugaverð lesning.

³ Consultative committee for space data systems. *Reference model for an open archival information system*. 2002. <http://public.ccsds.org/publications/archive/650x0b1.pdf> Bls. 4-1.

⁴ RLG. *Trusted digital repositories*. 2005. <http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf>

3.2 Virk notkun efnis

Í þessari skýrslu er fyrsti hluti lífsskeiðs safnefnis, upp að afhendingu þess til varðveislustofnunar, nefndur „virk notkun efnis“. Þessi hluti er nefndur hér þar sem huga þarf að varðveislustofnun er alltaf myndun þess, en ekki einungis frá afhendingu. Efni sem varðveita á, verður að jafnaði til utan varðveislustofnunar, hvort sem um skjalasafn eða bókasafn er að ræða. Skjalasöfn geta almennt haft áhrif á ferli við tilurð skjala, meðan bókasöfn eru mun háðari velvild efnismyndara.

Á Íslandi hefur fyrirkomulagið á varðveislustofnun opinberra gagna verið þannig, að varðveislustofnun fær almennt ekki efnið í sínar hendur fyrr en efnið er orðið óvirkt. Lög hafa skilgreint hvenær slíkt er og tilgreina ákveðinn árafjölda sem miða á við. Taka skal fram að þetta ferli á við allar tegundir efnis, líka stafrænt. Það getur hins vegar auðveldað starf varðveislustofnunar til muna ef hún getur haft áhrif á meðferð efnisins áður en hún fær það afhent. Hægt er, t.d. með lagasetningu og útgáfu leiðbeininga, að aðstoða stofnanir í að þróa vinnuáferðir og ferla sem stuðla að auðveldari varðveislustofnun seinna meir. Þetta er hægt ef varðveislustofnun er búin að skilgreina varðveislufærli hjá sér og hvaða kröfur þurfa að vera uppfylltar þegar efni er afhent henni, t.d. kröfur um efnissnið, lýsigögn og afhendingarferli.

Það er ljóst að með því að huga að þessum þætti í varðveislufærli, hvort sem um er að ræða skjalasöfn eða bókasöfn, er hægt að spara mikla vinnu. Þar sem skilgreint umfang skýrslunnar lýtur að safnefni eftir afhendingu, verður hins vegar ekki farið nánar út í hvaða aðgerða er þörf í þessum hluta.

3.3 Afhending/inntaka stafræns efnis

Efni er í flestum tilvikum undirbúið fyrir afhendingu af efnismyndara; vefvarsla er þó undantekning á þessu. Varðveislustofnun tekur við afhendingunni og skráir inn efnið. Þau atriði sem huga skal að hér í ferlinu eru:

- **Hversu ört og hvenær skila á efni.** Á Íslandi er skilgreint í lögum hvenær skila skal efni til varðveislustofnunar, en hægt er að fá undanþágu frá þessu með samkomulagi við viðkomandi varðveislustofnun.
- **Gagnasnið (e. format).** Því færri snið sem eru notuð, því betra. Snið henta auk þess misvel til varðveislustofnunar og almennt má segja að opin, einföld, stöðug (e. stable) og stöðluð snið henti best. (Sjá nánar um gagnasnið í kafla 5.3).
- **Lýsigögn (e. metadata).** Því betri sem lýsigögnin eru, því meiri líkur eru á að varðveitt skjöl verði trúverðug og nýtist í framtíðinni. Sjá nánar í kafla 3.5.
- **Miðill (e. medium).** Afhending á efni getur t.d. farið fram yfir net (vef eða tölvupóst), á segulmiðlum eða geisladiskum.
- **Dulkóðun (e. encryption).** Dulkóðun er talin fara illa saman við stafræna varðveislustofnun, þar sem hún eykur líkur á að skjal verði ólæsilegt í framtíðinni. Þó kemur til greina að nota dulkóðun við afhendingu, ef hún fer fram yfir opinn og óruggan miðil.
- **Tékksummur (e. check-sums).** Tékksummur eru notaðar til að bera saman skjöl fyrir og eftir afhendingu til að sannreyna að skjal hafi ekki breyst í ferlinu. Þá eru bitarnir í stafrænu skjölunum, þ.e. upprunalegu skjali og því skjali sem varðveislustofnun hefur fengið, lagðir saman og ef sama summa fæst er álitid að afhending hafi tekist. Tékksummur eru nokkurs konar lágmarksvörn gegn skjalabreytingum, en fremur auðvelt er að breyta skjölum án þess að breyta tékksummum. Rafrænar undirskriftir eru því enn betri leið til að sannreyna skil. Eðlilegt er að krafa um tékksummur eða rafrænar undirskriftir sé hluti af kröfum um stafræna afhendingu.

- **Rafrænar undirskriftir (e. digital signatures).** Rafrænar undirskriftir eru verulega frábrugðnar hefðbundnum undirskriftum. Í einfaldri mynd byggja þær á ákveðinni reikniaðferð sem, líkt og tékksummur, dregur efni skjals niður í lítið gagnamagn (sem þó er venjulega breytilegt eftir stærð skjals). Sé það gert fyrir og eftir afhendingu má sannreyna að skjalið sé óbreytt. Ólíkt tékksummum byggja rafrænar undirskriftir á dulkóðun og því er mjög erfitt að falska þær. Þannig breytist rafræn undirskrift nánast örugglega ef skjalinu er breytt á einhvern hátt, hvort sem það er fyrir slysi eða viljandi.

Rafrænar undirskriftir eru oft notaðar við flutning efnis til að votta gildi gagnanna (stundum er þeim beitt á tékksummur í stað skjalsins sjálfs, en það er ekki jafn örugg aðferð). Almennt eru rafrænar undirskriftir ekki geymdar til langs tíma af tveimur ástæðum. Í fyrsta lagi er ekki ljóst að hægt verði að sannreyna rafrænar undirskriftir í framtíðinni. Í öðru lagi njóta varðveislustofnanir trausts og varðveisla þeirra er talin næg trygging á því að efnið sé eins og það kom frá stofnunum. Þeim er því hent eftir gagnaflutning hjá þeim söfnum sem rætt var við í gerð þessarar skýrslu, en skrá má lýsigögn sem staðfesta tilvist rafrænna undirskrifa við afhendingu.

Almennt má segja að helsti munur á afhendingu opnins og lokaðs efnis er að skjalamyndarar bera almennt meiri ábyrgð á afhendingu varðandi lokað efni. Hér geta lagasetning og reglur stutt varðveislustofnanir við að setja fram kröfur um efnissnið, lýsigögn og afhendingarferli.

3.4 Örugg geymsla og varðveisla stafræns efnis

Að inntöku lokinni er efninu komið til varðveislu í tölvukerfi safnsins. Hér koma öryggismál til sögunnar sem snúa að aðgengi að efninu og árásum utan frá. Almennt má segja að helsti munur á opnu og lokuðu efni hvað varðar örugga geymslu og varðveislu felist í öryggisráðstöfunum. Þannig getur verið nægilegt að fyrir opið efni sé til traust afrit, því það er enginn stór skaði ef brotist er inn í tölvukerfi safnsins, viðkomandi kæmist ekki í neitt nema það sem aðgangur var að hvort eð er. Lokað efni kallar á meiri öryggisráðstafanir þar sem það er alvarlegt mál ef brotist er inn í tölvukerfi með lokuðu efni og viðkomandi kemst í persónulegt efni eða efni er varðar þjóðaröryggi.

Nokkur atriði þarf að hafa í huga fyrir örugga geymslu og varðveislu stafræns efnis, þau eru m.a.:

- **Varðveislusnið (e. preservation format).** Í þeim tilfellum sem skjölin berast safninu á ótryggu sniði þarf að færa þau á varðveislusnið. Eins og áður sagði eru opin, stöðluð snið talin henta best sem varðveislusnið.
- **Geymslutækni (e. storage technology).** Margir kjósa að geyma efni í vensluðum gagnagrunnum sem komin er mikil reynsla á hvað varðar öryggi gagna. Aðrir kjósa einfaldlega að geyma efni beint í skráarkerfi stýrikerfisins, en komin er löng reynsla á margar útfærslur skráarkerfa. Þriðji kosturinn sameinar þetta tvennt, það er að geyma efni í skráarkerfi sem útfært er ofan á vensluðum gagnagrunni. Í sumum tilfellum er efni geymt á tveimur mismunandi tölvukerfum til að takmarka líkur á að bilanir valdi gagnatapi.
- **Geymslumiðill (e. storage medium).** Flestir velja að geyma stafrænt efni á segulmiðlum, gjarnan samblöndu af diskastæðum og bandstöðvum. Þetta fyrirkomulag er gjarnan kallað „Hierarchical Storage Management“.⁵

⁵ Atkinson, S. *Storage management and backup schemes for broadcast video*. 2004. http://www.sgluk.com/downloads/docs/storage_management_lo_lo.pdf

Í einhverjum tilfellum eru notaðir aðrir miðlar, svo sem geisladiskar. Rökin fyrir valinu á geisladiskum eru tvenns konar. Í fyrsta lagi eru þeir fremur traustur varðveislumiðill til nokkurra ára ef rétt er farið með diskana. Þá er reyndar aðeins átt við vel framleidda geisladiska, en framleiðendur munu vera mjög misgóðir og nauðsynlegt að fylgjast mjög vel með gæðum diskanna. Hins vegar er mikil vinna að gera það svo vel sé.⁶ Í öðru lagi eru geisladiskar annars eðlis en segulmiðlar, og því ekki viðkvæmir fyrir segulmagni í umhverfinu. Þannig gætu hryðjuverkamenn mögulega valdið skemmdum á segulmiðlum og tölvubúnaði með framköllun á „rafsegulpúlssi“ (e. electro-magnetic pulse, EMP) en slíkar árásir hafa engin áhrif á geisladiska.⁷ Þess má þó geta í þessu samhengi að bandaríski herinn hefur unnið að því að efla varnir gegn rafsegulpúlsum í sínum tölvubúnaði og er hægt að verja segulmiðla með því að loka þá inni í stálbúrum, svokölluðum gauss-búrum.⁸

Brown (2003)⁹ sýndi fram á að geisladiskar séu mun öruggari geymslumiðill en harðir diskar. Hins vegar miðaðist sú umfjöllun við einstaka diska, en ekki þær diskastæður sem nú eru algengar, s.s. RAID stæður sem verja efni með því að skrá umframupplýsingar (e. redundant information) og geta skipt út diskum um leið og einn þeirra bilar. Bradley (2006)¹⁰ leiðir því að því rökum að diskastæður séu í raun hagkvæmari og öruggari miðill en geisladiskar, auk þess sem þær ráða augljóslega miklu betur við það mikla gagnamagn sem vista þarf nú til dags.

Loks má geta þess að Palm (2006)¹¹ stingur upp á því að tölvugerðar örfilmur (e. computer output microfilm, COM) geti verið öruggasti og ódýrasti langtímavarðveislumiðillinn, þar sem þær endast í allt að hundrað ár og eru auðlesnar. Í framtíðinni telur Palm að auðvelt verði að skanna efni þeirra inn og gera aðgengilegt (væntanlega þá á sama hátt og efni á segulböndum er núna gert aðgengilegt, án þess að það sé tekið fram). Hins vegar er spurning hvort ekki verði sama vandamál með vél- og hugbúnað fyrir þessar örfilmur, eins og aðra miðla, þegar fram í sækir.

- **Öryggisráðstafanir (e. security).** Segja má að lágmarksöryggi felist í því að geyma efni bak við eldvegg, þar sem virk öryggisstjórnun fer fram. Með mjög viðkvæmt efni er þó ástæða til að ganga skrefinu lengra. Tvær leiðir hafa verið farnar í því samhengi.

Mörg söfn (sér í lagi skjalasöfn) hafa þau tölvukerfi, sem geyma viðkvæmasta efnið, ótengd umheiminum og viðhafa sérstakar öryggisráðstafanir til að stýra aðgengi að þeim. Þá er það efni, sem veita á almennan aðgang að, flutt yfir á leitarkerfi sem eru aðgengileg umheiminum, en annað efni er afgreitt af starfsfólki safnsins. Þessi leið er tiltölulega ódýr í framkvæmd, miðað við hinn valkostinn sem ræddur er að neðan, og eins örugg og önnur öryggiskerfi safnsins.

Hin leiðin er að búa til mjög öruggt umhverfi með nýjustu tækni og tilheyrandi tilkostnaði. Þessi leið hefur t.d. verið farin í Bretlandi, en þar er skilgreint öruggt innra net stjórnsýslunnar, eins og fram hefur komið. Því til viðbótar notar National Archives

⁶ Statens arkiver. *Hvad har kornblomster med CD-R og elektronisk opbevaring af arkivalier at gøre?* 1999. <http://www.sa.dk/sa/itogarkiv/teknologi/CDRartikel/ARKIVCD.htm>

⁷ Dunn, J.R. *The EMP threat.* 2006. Skoðað 13.11.2006.

<http://www.leatherneck.com/forums/archive/index.php/t-28864.html>

⁸ Spencer, J. *The electromagnetic pulse commission warns of an old threat with a new face.* 2004. Skoðað 13.11.2006 <http://www.heritage.org/research/nationalsecurity/bg1784.cfm>

⁹ Brown, A. *Selecting storage media for long-term preservation.* 2003.

http://www.nationalarchives.gov.uk/documents/selecting_storage_media.pdf

¹⁰ Bradley, K. *Risks associated with the use of recordable CDs and DVDs as reliable storage media in archival collections.* 2006. <http://unesdoc.unesco.org/images/0014/001477/147782E.pdf>

¹¹ Palm, J. *The digital black hole.* 2006? http://www.tape-online.net/docs/palm_black_hole.pdf

eigin aðgangsstjórnun að efninu, til að geta stýrt aðgangi að einstöku efni. Þrátt fyrir að þetta net sé mjög öruggt, þá er ekkert 100% öruggt. Þá má geta þess að uppsetning og rekstur þessa nets kostar um 7,5 milljarða króna yfir 5 ára tímabil, en hugsanlega eru þó til ódýrari tæknilegar lausnir í dag.

Þótt netið, sem kerfið er á, sé mjög öruggt er full ástæða til að viðhafa aðgangsstjórnun á efninu sjálfu. Þar bjóða stýrikerfi og gagnasafnskerfi upp á öfluga valkosti, sá besti er líklega aðgangsstýring Oracle. Þar er gagnagrunnsstjóra leyft að bæta skilyrðum sjálfvirkri við hverja fyrirspurn einstakra notenda. Með réttum skilyrðum má tryggja að notendur fái aðeins aðgang að efni sem þeir hafa rétt til að sjá.

- **Afritun (e. back-up).** Afritun er ávallt nauðsynleg og mjög mikilvægt er að prófa afrit til að kanna hvort þau virki. Í mörgum tilvikum eru afrit tekin ört, t.d. daglega. Ef umsýslukerfið færir stafræna efnið aldrei til á diskum eða breytir því á annan hátt, er hins vegar mögulegt taka afrit sjaldnar, til dæmis þegar einungis efnið breytist. Þegar afrit eru tekin sjaldan, eru prófanir á afritum að sjálfsögðu ennþá mikilvægari.
- **Prófanir (e. testing).** Þegar treyst er á afrit sem varðveislutæki þarf að athuga þau reglulega með prófunum. Efnið er í flestum tilfellum haft á lifandi kerfi, þar sem hægt er að gera prófanir til að meta ástand þess. Auk þess að prófa efnið sjálft, þarf að prófa varðveislumiðlana og ástand þeirra.
- **Varðveisluafrit (e. preservation copies).** Þegar kemur að því hvernig varðveita á efnið, velja mörg söfn að geyma fleiri en eitt afrit af hverju skjali. Algengt er að efni sé bæði varðveitt á upprunalegu formi og á breyttu formi, svokölluðu varðveislusniði. Sum söfn hafa auk þess valið að búa til svokallað sýniafrit í stað þess að veita aðgang að varðveisluafritinu. Oftast er þetta gert með aðgangsstýringar í huga og til að auka öryggi varðveislueintaksins, sem er þá haldið sér. Stundum er sýniafritum breytt lítillega til að auka meðfærileika þeirra í leitarkerfum.
- **Langtímarvarðveisla (e. long-term preservation).** Þrátt fyrir að skjöl séu geymd á varðveislusniði, geta þau snið úrelst. Þá þarf annað hvort að færa skjölin á annað snið (e. migration) eða búa til búnað til að lesa sniðið sem hermír eftir upprunalega hug- og vélbúnaðinum (e. emulation).

Gagnaflutningur (e. migration) er þegar efni er flutt af einu kerfi yfir á annað eða af einu sniði yfir á annað. Þessi aðferð er ætluð til að koma til móts við hina öru tækniþróun og viðhalda aðgengi að efninu. Hér þarf að huga sérstaklega að gagnasniði, hvaða snið henta best til flutnings og lifa lengst. Það sem styður við að nota gagnaflutning sem aðferðafræði til langtímarvarðveislu, er að notendur geta skoðað efnið í kerfum og umhverfum sem þeir þekkja. Efnið er alltaf aðgengilegt með nýjustu tækni. Helsti gallinn við þessa aðferð er að við að færa efnið á annað snið/kerfi þarf að breyta því, sem getur orsakað villur í skjölunum. Auk þess geta villur hlaðist upp við fleiri gagnaflutninga og gæði efnisins rýrnað.

Hermun (e. emulation) er þegar vél- og/eða hugbúnaður er útbúinn til að herma eftir upprunalegum búnaði. Hermun er hugsuð sem aðferð til að hægt sé að viðhalda aðgengi að efninu í sínu upprunalega útliti og umhverfi. Þessi aðferð hentar betur en gagnaflutningur þegar um er að ræða flókið efni, þar sem útlit efnis og virkni getur tapast við flutning. Þetta á sérstaklega við margmiðlunarefni. Taka skal fram að hermun getur verið mjög dýr kostur og erfitt getur reynst að viðhalda henni til lengri tíma (jafnerfitt eða erfiðara en að viðhalda efninu).

Mikil umræða hefur farið fram um það hvor kosturinn sé betri en ýmsir telja að blanda af þessum aðferðum muni verða nauðsynleg. Öllum þykir ljóst að reglulega þarf að

skipta um vélbúnað, vegna þess hversu hratt hann úreldist. Einnig er talið líklegt að stundum þurfi að skipta um gagnasnið, ef engin leið er til að styðja ákveðið snið lengur. Þetta er hvort tveggja ákveðin tegund gagnaflutnings. Í öðrum tilfellum, þegar hugbúnaður úreldist eða skipt er um vélbúnað, getur þurft að skrifa eða kaupa nýjan hugbúnað. Þetta er ákveðin tegund hermunar þótt ekki sé endilega um að ræða allra fyrstu útgáfu efnisins. Ekki er enn ljóst hver besta blandan er eða hvort hægt verður yfirhöfuð að gefa eina forskrift fyrir langtíma varðveislu.

Varsla og varðveisla vefefnis verður alltaf vandkvæðum háð, því þetta efni notar oft kvika undirliggjandi tækni sem gerir það mjög erfitt að safna efninu og varðveita. Að auki er þetta efni á mjög fjölbreyttu sniði og ekki hægt að gera kröfur um að breyta því á varðveislusnið vegna mikils tilkostnaðar. Loks má minnast á varðveislu forrita ýmiss konar, en einhverjar erlendar rannsóknir hafa það að markmiði að varðveita forrit. Það er þó óhemju flókið viðfangsefni og getur ekki talist skynsamlegt að reyna slíka varðveislu í bráð.

3.5 Aðgangur og birting stafræns efnis

Söfnum ber lagaleg skylda að veita aðgengi að safnkosti sínum. Þar að auki er það álit skýrsluhöfunda að engin ástæða væri til að varðveita efnið ef ekki á að veita aðgengi að því. Nokkur álitamál eru til staðar þegar rætt er um aðgengi:

- **Hvernig er varðveislusniði breytt í birtingarsnið?** Í sumum tilvikum geta varðveislusnið og birtingarsnið verið sama sniðið, eins og t.d. getur gerst með myndefni. Í öðrum tilfellum þarf að fara fram vörpun í birtingarsnið, sem getur í einhverjum tilfellum þýtt að útlit skjalsins breytist í meðförum.
- Varpa þarf einfaldari sniðum, eins og t.d. XML og efni úr gagnagrunnum, yfir í birtingarform. Í tilfelli gagnagrunna er hægt að gera það á tvennan hátt, annars vegar með því að veita aðgang að einstökum færslum, eða hins vegar með því að flytja grunnana inn í gagnagrunnskerfi og vinna með þá þar. Fyrir seinni aðferðina getur þurft umtalsverða tölvuþekkingu til að framkvæma slíka vörpun.
- **Hvaða efni á almenningur að geta lesið?** Hér skiptir höfuðmáli hvort um er að ræða opið eða lokað efni. Lokað efni þarf að meðhöndla með varfærni, en það er spurning hversu „lokað“ það eru. Gæti almenningur t.d. haft aðgengi að lýsigögnum lokaðs efnis og þannig leitað í efninu? Eða er um að ræða algera lokun? Þetta getur verið snúið að ákveða.
 - **Hvaða efni á að vera aðgengilegt „utan safns“, t.d. yfir net?** Sum söfn setja opið efni sem ekki er varið höfundarrétti út á netið, t.d. lagasöfn. Önnur söfn velja að leyfa aðeins aðgang innan veggja safnsins. Höfundaréttarlögin setja hér nokkuð fastar skorður þar sem þau leyfa einungis meðferð stafræns efnis á safninu sjálftu.
 - **Hvernig geta notendur sannfært sig um að skjölin séu óbreytt, ófölsuð, áreiðanleg, og upprunaleg?** Þarna gegna lýsigögn höfuðhlutverki. Án lýsigagna, sem lýsa öllu ferlinu frá tilurð, gegnum afhendingu og varðveislu, og til birtingar, geta skjölin verið alfarið marklaus. Eins og minnst var á í kafla 3.3 eru tékksummur og rafrænar undirskriftir notaðar til að tryggja að efnið og lýsigögnin berist óbreytt frá skjalamyndara.

Lýsigögn (e. metadata) eru skilgreind sem gögn um efni, þ.e. þau lýsa efninu. Lýsigögn fyrir prentað efni lýsa hvað efnið inniheldur, hver skráði og hvenær, og einnig hvar það er vistað í safni. Lýsigögn fyrir stafrænt efni innihalda almennt séð sömu lýsigögnin og prentað efni, en þar sem mun auðveldara er að fylgjast með breytingum á stafrænu efni er almennt skráð meira af lýsigögnum. Lýsigögn stafræns

efnis skipta miklu máli og það má nánast segja að án lýsigagna sé stafrænt efni gagnslaust. Wikipedia¹² skilgreinir þrenns konar lýsigögn fyrir stafrænt safn:

1. Lýsingargögn: Lýsa innihaldi efnis líkt og er gert í safnaskrá. Hlutverk þessara lýsigagna snýr oftast að leit og endurheimt efnis.

Textaleitarvélar, eins og Google, eru orðnar mjög öflugar og sífellt er verið að rannsaka aðferðir til að leita eftir innihaldi mynda, hljóðs og annarra sniða. Í mörgum tilfellum er mun auðveldara að finna efni eftir innihaldi þess en eftir lýsigögnum, þótt lýsigögn eins og skráarstærð eða tegund geti einnig hjálpað til.

2. Uppbyggingargögn: Lýsigögn sem tengja saman allt það efni sem myndar eina heild.
3. Umsýslugögn: Upplýsingar sem notaðar eru til að stýra aðgangi að efni, t.d. upplýsingar um varðveislusnið og höfundarrétt, og upplýsingar um feril gagnanna, t.d. í langtímarvarðveislu.

Í gagnalýsingasafni PREMIS (PREservation Metadata: Implementation Strategies) hópsins fyrir varðveislulýsigögn,¹³ er lýsigögnum fyrir langtímarvarðveislu skipt í fimm þætti, sem snúa að eignarhaldi (e. provenance, hver hefur varðveitt efnið), upprunaleika (e. authenticity, er skjalið það sem það segist vera), varðveisluáðgerðum (e. preservation activity, hvernig hefur efninu verið breytt síðan það kom til varðveislu), tækniumhverfis (e. technical environment, hvað þarf til að hægt sé að birta efnið), og réttindamála (e. rights management, hver á skjalið og hver má sjá það).

Þegar rætt er um aðgengi, má segja að flokkun efnis í opið eða lokað efni skipti sköpum. Lagaumhverfið setur fastar línur og hefur mikil áhrif á aðgengi að efninu. Sterk persónuverndarlög hljóta að kalla á mikla aðgæslu með aðgangi að slíku efni, sérstaklega ef veita á beinan aðgang að safnefni.

¹² <http://en.wikipedia.org/wiki/Metadata>

¹³ PREMIS. *Data dictionary for preservation metadata*. 2005. <http://www.oclc.org/research/projects/pmwg/premis-final.pdf>

4 Fordæmi

Mikið af almennu efni um stöðu verkefna innan langtímavarðveislu var skoðað, auk þess sem rætt var við starfsfólk safna á Íslandi. Eftir þessa yfirferð voru nokkur söfn valin sem til þess fallin að skoða nánar.

Alþjóðasamband bókasafna- og bókavarðafélaga (IFLA) gaf út könnun á árinu þar sem könnuð var staða langtímavarðveislu stafrænna gagna.¹⁴ Samkvæmt þeirri skýrslu voru þau landsbókasöfn sem lengst voru komin í heiminum í varðveislu stafræns efnis, þjóðbókasöfnin í Hollandi og Ástralíu, og því voru þau skoðuð nánar. Þjóðbókasafn Hollands hefur verið leiðandi í langtímavarðveislu tímarita í samstarfi við fjölmarga alþjóðlega útgefendur. Þjóðbókasafn Ástralíu hefur einnig verið virkt, m.a. í vistun vefgagna gegnum PANDORA verkefnið. Í öðrum löndum var stöðunni almennt lýst þannig að þau væru meðvituð um þörf á langtímavarðveislu, væru að hefja störf, en ekki langt komin. Nú er ljóst að mörg söfn eru að sinna spennandi verkefnum á þessu sviði, til dæmis eru mörg háskólabókasöfn í Bandaríkjunum að vinna að langtímavarðveislu, en vegna takmarkaðs umfangs skýrslunnar reyndist ekki unnt að skoða fleiri bókasöfn.

Í ljósi þess að eðli stafræns efnis þjóðskjalasafna og aðgengis að því gerir varðveislu að sumu leyti erfiðari, og þess að mörg athyglisverð skjalasöfn fundust við leit, voru fleiri þjóðskjalasöfn skoðuð nánar. Þau lönd sem ákveðið var að skoða eru: Danmörk, Svíþjóð, Holland, Ástralía, Bretland og Bandaríkin. Svíþjóð hefur mótað athyglisvert rannsóknarverkefni um langtímavarðveislu og var fróðlegt að skoða áherslur þeirra. Danmörk varð fyrir valinu vegna áherslu þjóðskjalasafns Danmerkur á varðveislu stafræns efnis og samvinnu við Þjóðskjalasafn Íslands. Þjóðskjalasafn Hollands hefur verið í samstarfi við þjóðbókasafn Hollands. Þjóðskjalasafn Ástralíu hefur þegar komið upp varðveislukerfi. Þjóðskjalasafn Bretlands er með verðlaunað kerfi og hjá NARA í Bandaríkjunum er unnið að því að þróa kerfi byggt að hluta á sama búnaði og hjá Bretlandi. Mörg önnur lönd hafa komið upp við gagnaöflun en voru ekki skoðuð ítarlega, þar sem tíminn var takmarkaður.

Loks var Kanada skoðað, en þar hefur starfsemi þjóðskjalasafns og þjóðbókasafns verið sameinuð.

Lesið var tiltækt efni á vefnum frá þeim söfnum sem skoðuð voru og spurningar voru auk þess sendar til fulltrúa hinna erlendu safna.

4.1 Ísland

Á Íslandi er skipting efnis milli Þjóðskjalasafns Íslands (ÞÍ) og Landsbókasafns Íslands – Háskólabókasafns (Lbs-Hbs) nokkuð hefðbundin. Lbs-Hbs tekur við öllu útgefnu efni, þ.m.t. lögum og annarri útgáfu frá ríkinu ásamt einkaskjalasöfnum í handritadeild. Þjóðskjalasafn tekur við öllu opinberu efni ásamt ýmsum einkaskjalasöfnum. Héraðsskjalasöfn taka við sveitarstjórnarskjölum og einkaskjölum úr héraði.

Í undirköflum hér að neðan er farið yfir stöðuna hjá söfnunum, samkvæmt upplýsingum starfsfólks þeirra. Skýrsluhöfundum virðist ljóst að bæði söfnin hafi búið við mjög þröngan kost hvað varðar fjármagn til undirbúnings og framkvæmdar langtímavarðveislu stafræns efnis og raunar hefðbundins efnis einnig. Söfnin hafa þrátt fyrir það starfað um allnokkurn tíma að undirbúningi langtímavarðveislu, og hafa tekið þátt í ýmsu erlendu samstarfi á því sviði. Þau eru þó, eins og flest önnur söfn í heiminum, komin skammt á veg í tæknilegri úrlausn á langtímavarðveislu.

¹⁴ Verheul, I. *Networking for digital preservation*. 2006.

Ýmis önnur söfn og stofnanir varðveita, eða munu varðveita, stafrænt efni, s.s. Kvikmyndasafn Íslands og RÚV. Ekki er vitað til að þessir aðilar hafi hugað að langtímarvarðveislu, en hins vegar ættu lausnir annarra safna að geta nýst þeim.

4.1.1 Lagaumhverfi

Í þessum kafla verður stiklað á stóru á þeim helstu lagaákvæðum á Íslandi sem eru mikilvæg þegar kemur að langtímarvarðveislu stafræns efnis. Er þetta gert til að sýna hvaða kröfur núverandi lagaumhverfi gerir til þessa verkefnis.

Stjórnvöld á Íslandi búa við víðtæka upplýsingaskyldu. Á móti þessari upplýsingaskyldu stjórnvalda kemur réttur einstaklingsins og önnur atriði, eins og þjóðaröryggi. Þessi atriði takmarka upplýsingaskylduna og vernda viðkvæmar upplýsingar. Þessir tveir pólar eru sífellt að takast á og hefur verið reynt að búa til leiðbeiningar í formi lagasetningar til að gera það ljóst hvernig meðferð þessa efnis skal háttað.

Í skýrslu nefndar um rafræna stjórnsýslu frá 2002, er rætt um upplýsingaskyldu stjórnvalda: „Samkvæmt 2. tölulið 2. mgr. 3. gr. upplýsingalaga nr. 50/1996 ber stjórnvöldum ekki aðeins skylda til að veita almenningi aðgang að skjölum máls, heldur einnig að öllum öðrum gögnum sem mál varða, svo sem teikningum, uppdráttum, kortum, myndum, örfilmum og gögnum, sem vistuð eru í tölvu. Samkvæmt þessu taka upplýsingalög ótvírætt til rafrænna gagna.“¹⁵

Lögin segja líka til um hvers konar efni skal berast stofnunum, þó er einkaaðilum velkomið að gefa efni til safnanna þar fyrir utan.

Þjóðskjalasafn Íslands varðveitir efni frá opinberum stofnunum og eru það aðallega skjöl sem berast safninu. Efni Landsbókasafns Íslands – Háskólabókasafns berst aðallega í gegnum skylduskil.¹⁶ Lög um skylduskil¹⁶ til safna skilgreina Landsbókasafn Íslands – Háskólabókasafn, Amtsbókasafnið á Akureyri og Kvikmyndasafn Íslands sem varðveislusöfn skyldueintaka. Í lögnum segir að „af verkum sem gefin eru út eða birt hér á landi skal afhenda eintök til skylduskila ...Verk telst birt í skilningi laga þessara þegar eintök af því hafa verið löglega gefin út eða þegar það er með rétttri heimild gert aðgengilegt almenningi með öðrum hætti, svo sem ef það er flutt eða sýnt opinberlega eða er aðgengilegt almenningi á rafrænu formi um tölvunet. Verk telst birt hér á landi þótt það sé framleitt erlendis ef það er sérstaklega ætlað til dreifingar hér á landi.“

Efni skal skila í endanlegri gerð og á því formi sem það birtist notendum. Safninu er einnig heimilt að semja við skilaskyldan aðila um sérstakan frágang á efni ef þurfa þykir.

Þar sem þessi skýrsla er um stafrænt efni, er vert að velta fyrir sér hvernig lögin líta á efni á stafrænu formi. Í stjórnsýslulögnum¹⁷ er kafla um rafræna stjórnsýslu og hvaða kröfur eru gerðar, bæði til stjórnvalda og notenda. Þessi ákvæði eiga við um þegar stjórnvöld taka ákvarðanir í einstökum málum varðandi réttindi og skyldur borgaranna. Í 36. gr. segir t.d. að efni á rafrænu formi telst fullnægjandi sé form þannig að það sé tæknilega aðgengilegt móttakanda og hann geti kynnt sér efni þess, varðveitt það og framvísað því síðar. Einnig er tekið fram í 37. gr. að rafrænt frumrit er talið fullnægjandi sé það tryggt að það sé óbreytt frá upprunalegri gerð. Rafrænar undirskriftir geta verið verkfæri sem notað er til að tryggja þessi lagaákvæði og 38. gr. stjórnsýslulaganna fjallar um þær. Þar segir að rafræn undirskrift getur komið í stað eiginhandarundirskriftar, „enda tryggi rafræna undirskriftin með sambærilegum

¹⁵ Forsætisráðuneytið. *Skýrsla nefndar um rafræna stjórnsýslu*. 2002. Bls. 25

¹⁶ *Lög um skylduskil til safna nr. 20/2002*.

¹⁷ *Stjórnsýslulög nr. 37/1993*.

hætti og eiginhandarundirskrift persónulega staðfestingu þess sem gögnin stafa frá. Fullgild rafræn undirskrift samkvæmt lögum um rafrænar undirskriftir skal ætíð teljast fullnægja áskilnaði laga um undirskrift.“

Stjórnsýslulögin gera þær kröfur til stafræns efnis að það geti verið aðgengilegt móttakendum og aðilum máls, og að hægt sé að tryggja með afgerandi hætti að það sé óbreytt og í upprunalegu formi. Um varðveislu efnisins segir í 40.gr. að stjórnvald „skal varðveita rafræn gögn þannig að unnt sé að sannreyna efni og uppruna þeirra síðar með aðgengilegum hætti.“

Hinn víðtæki upplýsingaréttur almennings kallar á að aðgengi að efninu sé þannig að hægt sé að verða við beiðni um aðgang með skjótum hætti. Upplýsingalögin voru sett í maí 1996¹⁸ og rýmkuðu til muna aðgang almennings að gögnum og upplýsingum hjá stjórnvöldum. Samtímis setja þau upplýsingaréttinum skorður um efni sem varðar m.a. almannaheill og einkahagsmuni. Þessi skilyrði setja starfsmönnum Þjóðskjalasafns Íslands ákveðnar takmarkanir þegar afgreitt er úr safninu og er mjög mikilvægt að rétt sé farið með efnið. Að þessu þarf því sérstaklega að huga þegar kemur að aðgengi að stafrænu efni. Lög um Þjóðskjalasafn Íslands nr. 66/1985 tilgreina að aðgangur að skjölum og öðru efni sem varðveitt er í safninu, fari samkvæmt ákvæðum upplýsingalaga. Um aðgang að öðru efni og skjölum, sem upplýsingalög taka ekki til, skal mælt fyrir um í reglugerð sem menntamálaráðherra setur að fengnum tillögum þjóðskjalavarðar.¹⁹

Þegar kemur að rétti einstaklingsins og verndun persónulegs efnis eru það upplýsingalögin og stjórnsýslulögin sem koma til sögu. 5. gr. upplýsingalaga segja: „Óheimilt er að veita almenningi aðgang að gögnum um einka- eða fjárhagsmálefni einstaklinga sem sanngjarnt er og eðlilegt að leynt fari, nema sá samþykki sem í hlut á. Sömu takmarkanir gilda um aðgang að gögnum er varða mikilvæga fjárhags- eða viðskiptahagsmuni fyrirtækja og annarra lögaðila.“ 15.-17.gr. stjórnsýslulaganna segja til um upplýsingarétt aðila máls og takmarkanir á honum. Þar er segir m.a. „Þegar sérstaklega stendur á er stjórnvaldi heimilt að takmarka aðgang aðila máls að gögnum ef hagsmunir hans af því að notfæra sér vitneskju úr þeim þykja eiga að víkja fyrir mun ríkari almanna- eða einkahagsmunum ...“.

Skv. upplýsingalögum er stjórnvöldum er skylt að veita aðilanum sjálfum aðgang að skjölum og öðru efni sem varðar tiltekið mál ef það hefur að geyma upplýsingar um hann sjálfan. Þessi aðgangur gildir alltaf, ekki einungis eftir að efnið er orðið 30 ára (sem er almenna viðmiðunarreglan), en þetta gildir þó ekki um efni sem fellur undir takmarkaðan aðgang eða efni sem geymir upplýsingar um mikilvæga almannahagsmuni er leynt eiga að fara. Auk þess er heimilt að takmarka aðgang aðila að efni ef það hefur jafnframt að geyma upplýsingar um einkamálefni annarra, enda vegi þeir hagsmunir, sem mæla með því að upplýsingunum sé haldið leyndum, þyngra en hagsmunir þess sem fer fram á aðgang að efninu.

Ef vafi er um rétt til aðgangs að efninu getur safnið aflað rökstuddrar umsagnar þess stjórnvalds sem afhenti safninu efnið áður en ákvörðun er tekin.

Um aðgengi að efni Lbs - Hbs segir í 13. gr. laga um skylduskil að það „efni sem er trúnaðar- eða einkamál og aðeins birt til mjög takmarkaðra afnota getur útgefandi ákveðið að móttökusafn geymi innsiglað tiltekinn tíma og er safni skylt að heita því með skriflegri yfirlýsingu ef óskað er. Sá tími sem efni er þannig innsiglað að ósk útgefanda má ekki fara fram úr gildistíma höfundaréttar. Útgefandi getur ákveðið að þann tíma sem prentað eða fjölfaldað efni er innsiglað séu öll skyldueintök varðveitt í Landsbókasafni. Af efni sem

¹⁸ *Upplýsingalög nr. 50/1996.*

¹⁹ Sjá hér t.d. úrskurð úrskurðarnefndar um upplýsingamál í máli nr. A-235/2006

opinber yfirvöld gera upptækt skulu lögreglu yfirvöld tryggja skil eintaka og kveða á um innsíglun þeirra.“

Þar að auki eru öll skilaskyld verk aðgengileg innan heimildamarka höfundalaganna²⁰. 3. gr. höfundalaganna segir frá rétti höfundar: „Höfundur hefur einkarétt til að gera eintök af verki sínu og til að birta það í upphaflegri mynd eða breyttri, í þýðingu og öðrum aðlögunum.“ Þetta þýðir að hafi Landsbókasafn-Háskólabókasafn rit á stafrænu formi, hefur safnið ekki heimild til að veita aðgang að því utan safns.

Almenna regla höfundaréttarlaganna er að réttur höfundar gildir í 70 ár eftir lát hans skv. 43. gr. laganna. Upplýsingalögin breyttu þessu aðeins á þann veg að heimilt er að veita aðgang að efni með því að afhenda ljósrit eða afrit af þeim, þó svo að efnið sé verndað af höfundaréttarlögum. Þessi heimild takmarkast við að efnið verði ekki birt, fjölfaldað og dreift nema með samþykki höfundar.

Þar að auki segja upplýsingalögin að veita skuli aðgang að efni, sem fellur ekki undir takmarkaðan aðgang, þegar liðin eru þrjátíu ár frá því að efni varð til, að frátöldum upplýsingum er varða einkamálefni einstaklinga, en aðgang að þeim skal fyrst veita að áttatíu árum liðnum frá því að þau urðu til.

Hér hefur verið farið gróflega yfir það lagaumhverfi sem ÞÍ og Lbs-Hbs búa við, en rétt er að hnykkja á ákveðnum atriðum sem standa upp úr eftir þessa umfjöllun:

- Stofnanir bera ábyrgð á aðgengi að sínum gögnum í 30 ár áður en þau eru afhent varðveislustofnun.
- Varðveislustofnanir þurfa að gera kröfur vegna afhendingar efnis.
- Viðkvæmt efni og efni sem fellur undir höfundalög setur varðveislustofnunum ákveðnar skorður.
- Ákveðin tímamörk gilda varðandi hvenær aðgengi að efni er opnað.
- Kröfur eru gerðar um tækni og öryggi og aðgengileika safnefnis.

Í kafla 6.2 verður fjallað nánar um þessi atriði og hvað þarf að hafa í huga til að takast á við þau.

4.1.2 Þjóðskjalasafn Íslands

Þjóðskjalasafn Íslands tekur við efni frá öllum opinberum stofnunum, undirstofnunum þeirra, ríkisreknum fyrirtækjum, félagasamtökum sem fá meiri hluta rekstrarfjár síns með framlagi á fjárlögum og félög sem njóta verulega styrks af opinberu fé. Alls eru þetta nærri þúsund aðilar.²¹ Skilaskyldar stofnanir skulu afhenda safninu skjölin sín „eigi síðar en þegar þau hafa náð þrjátíu ára aldri. Er þá miðað við síðustu innfærslu í bók eða síðasta bréf afgreidds máls. Þjóðskjalavörður getur lengt þennan frest eða stýtt í einstökum tilvikum ef sérstakar ástæður mæla með því.“²² Fram að þeim tíma ber hver stofnun því ábyrgð á skjölunum og heldur utan um aðgang að þeim.

²⁰ Höfundalög nr. 73/1972.

²¹ Þjóðskjalasafn Íslands, *Rafræn skjala- og gagnavarsla ríkisstofnana*. 2005. Bls. 6

²² Lög um Þjóðskjalasafn, nr. 66/1985, 6. gr.

Þegar kemur að afhendingu efnis til Þjóðskjalasafns, gerir safnið vissar kröfur um skráningu, flokkun og frágang skjala. Í leiðbeiningum safnsins um afhendingu skjala og gerð geymsluskráar er gerð ítarleg grein fyrir hvernig gengið skuli frá skjalasöfnum til afhendingar.

Þessi skilyrði um vistun og frágang skjala hjá stofnunum gera það að verkum að ákveðið verklag er haft á skjalavistun í mörgum stofnunum. Skilyrðin auðvelda auk þess safninu að taka við gögnum og veita aðgang að þeim, sé farið að þeim. Það er því miður ekki alltaf gert og glímir safnið við margra ára vanda af þeim sökum og vegna skorts á starfsfólki og húsnæði.

Sú grundvallarregla í skjalavörslu sem ÞÍ starfar eftir er svokölluð upprunaregla, sem felur í sér að varðveita efni í því samhengi sem það varð til í. Þannig er öllum afhendingum til ÞÍ haldið sem heildareiningum. Langtímavarðveisla stafræns efnis þarf að taka tillit til þessarar reglu.

Til að mæta kröfum nýrrar tækni og tilkomu stafræns efnis hefur safnið útbúið drög að reglum um rafræn gagna- og skjalasöfn opinberra aðila. Það hefur hins vegar komið fram í samtali við starfsfólk safnsins að best væri að þessar reglur yrðu gerðar að reglugerðarígildi, svo að safnið hafi enn sterkari stöðu lagalega gagnvart skilaskyldum stofnunum.

ÞÍ hefur hafið tæknilegan undirbúning að formlegri móttöku stafrænna skila. Tilraunaverkefni hefur verið sett í gang með Ríkisskattstjóra um skil á gagnagrunni virðisaukaskattkerfis. Einnig er í gangi tilraunaverkefni með menntamálaráðuneytinu þar sem verið er að skoða stafræn skil úr málaskrá ráðuneytisins til safnsins.

Árið 1997 var sett á fót nefnd á vegum menntamálaráðuneytisins um varðveislu tölvugagna sem verða til í stjórnsýslunni. Þessi nefnd komst að þeirri niðurstöðu að Ísland ætti að taka sér til fyrirmyndar aðferðir þjóðskjalasafns Danmerkur, Statens Arkiver, í þessum málum.²³ Að því hefur ÞÍ starfað síðan. Safnið hefur gert samning við Statens Arkiver um óheftan aðgang að öllum hugbúnaði þeirra og reynslu á sviði langtímavarðveislu, en ekki hefur enn verið fjárfest í búnaði til varðveislu stafræns efnis. Þau regludrög sem ÞÍ hefur lagt til um stafræn skil til safnsins eru þýdd og staðfærð úr dönsku reglunum með styrk úr Vísindasjóði í samstarfi við Hugvit. Að þessu marki vísast til kafla um Statens Arkiver hér að neðan um þá stefnu sem ÞÍ er að móta sér í langtímavarðveislu.

4.1.3 Landsbókasafn Íslands - Háskólabókasafn

Eins og kemur fram hér að framan, berst íslenskt efni til Landsbókasafns Íslands – Háskólabókasafns í samræmi við lög um skylduskil. Hlutverk safnsins er mjög viðamikilið og felur það m.a. í sér söfnun á útgefnu efni á Íslandi í hvaða formi sem er, þar með talið vefefni og tón- og myndefni, en að auki er safnað efni um Ísland eða Íslendinga sem útgefið er erlendis. Jafnframt er tekið á móti einkaskjalasöfnum og söfnum frá félagasamtökum. Í framtíðarsýn safnsins er gert ráð fyrir að safnið sé þekkingarveita og að almenningur fái aðgang að sem víðtækustum hluta safnkostsins gegnum eina gátt, þjóðbókagáttina.

Lbs-Hbs er með í undirbúningi að taka við efni á stafrænu formi, t.d. blöðum, bókum, rannsóknarritgerðum og efni úr munnlegri geymd. Íslensk tímarit fram til 1920, handrit og kort hafa verið færð í stafrænt form og unnið er að því að skanna inn íslensk dagblöð. Stafræna efnið er skráð í Gegni eins og annað efni auk þess sem aðgangur er að fjölbreyttu efni á stafrænu formi á vef safnsins, þó efnið sé ekki vistað í safninu sjálfu. Þar má t.d. nefna

²³ Menntamálaráðuneytið. *Skýrsla nefndar um varðveislu tölvugagna sem verða til í stjórnsýslunni*. 1998. Bls. 14

aðgang að gagnasöfnum og tímaritum. Þá hefur safnið um nokkurt skeið safnað íslensku vefefni með vefsöfnunarkerfi sem heitir Heritrix, en ekki er komin endanleg lausn á aðgengi að því efni.

Ekki hefur verið unnið formlega að stafrænni langtímarvarðveislu hjá safninu en það fylgist náið með því sem verið er að gera annars staðar. Í vinnslu er kynningarefni og leiðbeiningar til vefstjóra vegna vefsöfnunarinnar en til eru leiðbeiningar um skil á pappírsefni og handritum. Stafræna efnið er geymt á tveimur RAID diskastæðum og á segulböndum og í undirbúningi er að skoða og skilgreina þörf safnsins fyrir stafrænar gagnageymslur (e. trusted digital repository).

4.2 Ástralía

Skipting varðveisluhlutverks er hefðbundin milli National Library of Australia (NLA) og National Archives of Australia (NAA).

Þjóðskjalasafnið tekur við öllu efni sem er orðið 30 ára, en þá falla þau ekki lengur undir persónuverndarlögin þar í landi. Ef stofnun vill að efni eigi að haldast lokað eftir afhendingu skal gera ítarlega grein fyrir því; ekki er nóg að merkja það þannig. Þar af leiðandi getur starfsfólk þjóðskjalasafnsins opnað aðgang að efni þar sem ekki eru nægileg rök fyrir takmörkuðu aðgengi. Um þetta eru ákvæði í lögum um þjóðskjalasafn og ráða þau því miklu um aðgengi og er starfsfólk safnsins í aðstöðu til að meta efnið og hvernig aðgengi að því á að vera.²⁴

Þjóðbókasafninu ber skv. lögum um þjóðbókasafn að halda við og safna öllu sem við kemur Ástralíu og Áströlum.

4.2.1 National Archives of Australia (NAA)

Segja má að aðferðafræði NAA felist í þremur meginatriðum:

- Efni er fært á staðlað form sem er byggt á XML. Þar að auki er upprunalega efnið fært á staðlað textasnið sem kóðar tvíundarsnið efnisins (base64)²⁵ og er auðvelt að færa til baka á upprunalega sniðið; þetta textaskjal er svo vafið inn í XML. Þannig eru geymdar tvær útgáfur af efninu. Þetta seinna eintak er gert til að eiga upprunalegu útgáfuna, ef notendur seinni tíma eru svo lánsamir að eiga hugbúnað sem getur birt hana.
- Umfangsmikið ferli hefur verið skilgreint, og hugbúnaður skrifaður, til að færa efnið í örugga gagnageymslu til varðveislu. Þetta ferli fer fram á þremur einangruðum netum, og efni er flutt milli þeirra á hörðum diskum. Sérstakur hugbúnaðarpakki, Digital Preservation Recorder (DPR), sér um að fylgjast með ferlinu og færa lýsigögn á milli lokuðra netanna.
- Aðgangur er veittur gegnum RecordSearch, sem inniheldur öll lýsigögn og býður upp á leit yfir Internetið. Opið efni er fært á diskum yfir í RecordSearch og þannig gert aðgengilegt. Annað efni er birt notendum, þegar við á, á lokuðum vélum innan safns.

Gagnasniðið sem NAA hefur valið til langtímarvarðveislu hefur safnið sjálft unnið.²⁶ Í stuttu máli er þetta XML snið sem inniheldur kjarna (e. essence) efnisins.²⁷ Fyrir textaskjöl er t.d.

²⁴ National archives of Australia. *Records and the public*. Skoðað 29.10.2006.

http://www.naa.gov.au/recordkeeping/access/records_public.html

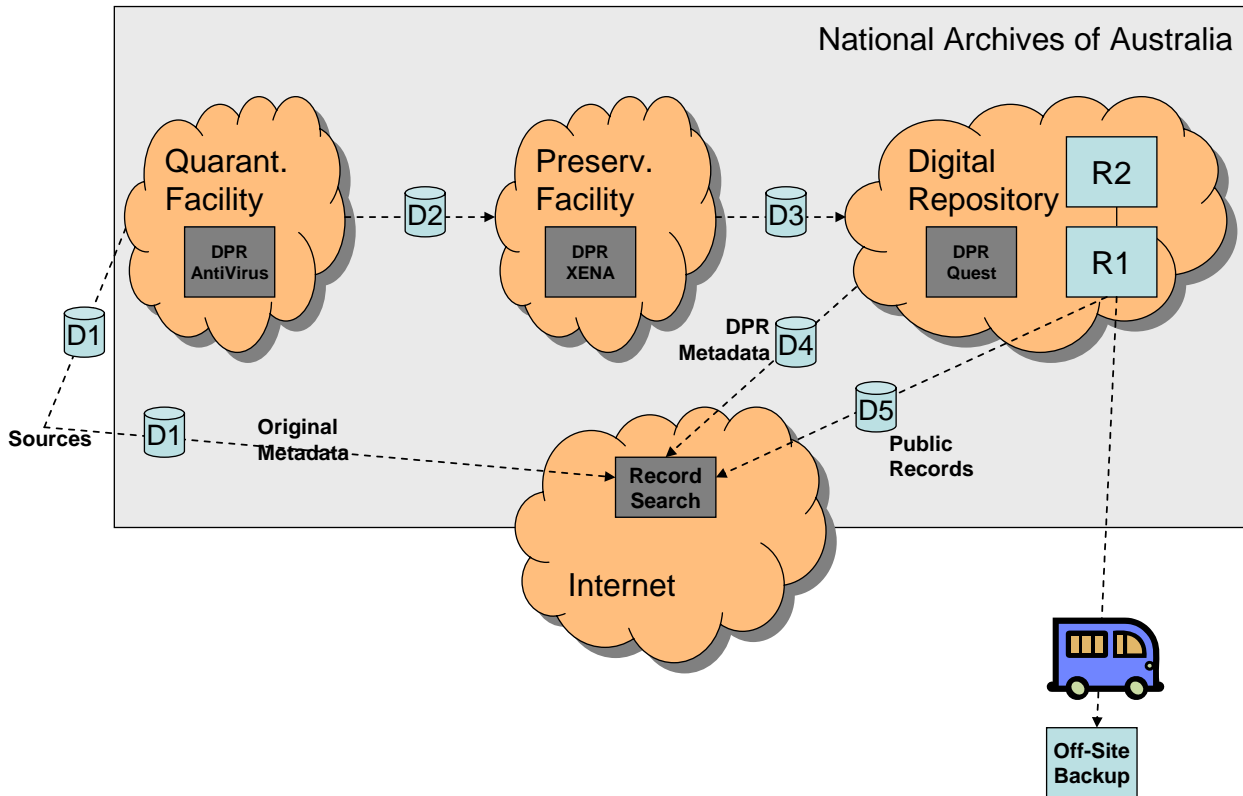
²⁵ Wikipedia. *Base64*. Skoðað 06.12.2006. <http://en.wikipedia.org/wiki/Base64>

²⁶ National archives of Australia. Skoðað 29.10.2006.

http://www.naa.gov.au/recordkeeping/preservation/digital/xml_data_formats.html

geymdur textinn ásamt uppbyggingu skjalsins. Myndum er varpað á textasnið með base64. Flókin skjöl eru skilgreind sem „pakki“.

Mynd 4-1 gefur yfirsýn yfir högun (e. architecture) varðveislu- og aðgangs hjá NAA, sjá einnig bæklinginn *Digital preservation: illuminating the past, guiding the future*,²⁸ og skýrslurnar *An approach to the preservation of digital records*,²⁹ *Digital preservation: Digital preservation software applications*³⁰ og *Digital preservation project*.³¹



Mynd 4-1 : Yfirlit yfir varðveisluferli hjá National Archives of Australia

Lýsigögn eru fyrst send inn frá stofnunum, og því næst er efni sent inn á hörðum disk, eða í tölvupósti ef um litlar sendingar er að ræða (D1). Lýsigögn eru gerð samkvæmt áströlskum staðli um efni í stjórnsýslunni, „The Commonwealth record series (CRS) system.“³²

Lýsigögnin eru geymd í RecordSearch kerfinu (sjá umfjöllun síðar), en efnið sjálft fer inn í umfangsmikið varðveisluferli sem er lýst hér á eftir.

Efnið er fyrst fært inn í svokallaða sóttkví, þar sem m.a. er athugað að tékksummur standist og efnið vírusleitað. Ef rafrænar undirskriftir eru til staðar er þeim hent á þessu stigi málsins.

²⁷ National archives of Australia. *An approach to preservation of digital records*. 2002.

http://www.naa.gov.au/recordkeeping/er/digital_preservation/Green_Paper.pdf

²⁸ National archives of Australia. *Digital preservation: illuminating the past, guiding the future*. 2006.

http://www.naa.gov.au/recordkeeping/preservation/digital/XENA_brochure.pdf

²⁹ National archives of Australia. *An approach to preservation of digital records*. 2002.

http://www.naa.gov.au/recordkeeping/er/digital_preservation/Green_Paper.pdf

³⁰ National archives of Australia. *Digital preservation: digital preservation software applications*. Skoðað 29.10.2006. <http://www.naa.gov.au/recordkeeping/preservation/digital/applications.html>

³¹ National archives of Australia. *Digital preservation project*. [An animation]. Skoðað 29.10.2006.

<http://www.naa.gov.au/recordkeeping/preservation/digital/animation/index.html>

³² National archives of Australia. *The Commonwealth record series (CRS) system*. Skoðað 29.10.2006.

<http://www.naa.gov.au/fsheets/fs06.html>

Digital Preservation Recorder pakkinn (sjá að ofan) skráir allar aðgerðir sem gerðar eru á efninu. Ef efnið er í lagi er það geymt á ótengdum hörðum diskum í u.þ.b. fjórar vikur og ferlið þá endurtekið, t.d. til að finna vírusa sem voru óþekktir á afhendingartíma. Ef efnið er enn í lagi að þeim tíma loknum, eru lýsigögn úr DPR sett á harðan disk með efninu og diskurinn síðan fluttur í varðveisluaðstöðuna (D2).

Í varðveisluaðstöðunni er aftur farið yfir tékksummur og heilleika efnis. Síðan eru búin til tvö eintök af efninu, með heimasmíðuðum hugbúnaðarpakka sem heitir XENA:³³

- Í fyrsta lagi er efninu varpað í staðlað textasnið sem kóðar tvíundargildi efnisins (base64).
- Í öðru lagi er kjarna efnisins varpað yfir á staðlað XML snið.³⁴ Fyrir skjöl á PDF sniði eru geymdar upplýsingar um hverja síðu, hvað er að finna á síðunni og hvar. Innihald getur verið skilgreint sem mynd annars vegar og texti hins vegar. Einnig er sérstakt snið fyrir gagnagrunna.

Bæði þessi eintök, ásamt lýsigögnum úr DPR, eru sett á enn einn harðan disk (D3) og flutt yfir í stafrænu geymsluna. Þar er efnið geymt á tveimur mismunandi gagnageymslum (R1 og R2). Þessar gagnageymslur nota mismunandi stýrikerfi og mismunandi diskageymslur (R1 = Dell 2600, Windows 2000, EMC Clariion RAID, Dell SLDT bandstöð; R2 = Dell 2600, Red Hat Linux, Apple Xserve RAID). Að auki eru afrit af R1 flutt yfir í aðra örugga geymslu.

Í framtíðinni er ætlunin að flytja aðra gagnageymsluna út úr húsi og nota ljósleiðaratengingu til að halda gagnageymslunum samræmdum. Efni sem er sérlega viðkvæmt verður þá flutt á öruggisbílum milli staða.

Kerfið sem heldur utan um efnið í gagnageymslunum heitir Quest. Það er heimasmíðað kerfi sem geymir vel valin lýsigögn sem nauðsynleg eru fyrir varðveisluna og til að finna efnið aftur. Quest er ekki notað til að leita að efni, heldur er fyrst og fremst notað til að tryggja aðgengi að því þegar þarf. Ef efni er sótt í Quest eru það fært á sérstakt útflutningssvæði og þaðan má afrita það á harðan disk til flutnings.

Aðgengi að lýsigögnum og opnu efni er yfir vef, gegnum kerfi sem heitir RecordSearch, sem er heimasmíðað kerfi sem byggir að hluta á SQL Server. Hluti lýsigagna kemur frá skjalamyndara (D1) en hluti kemur úr lýsigögnum þeim sem DPR safnar (D4).

Lýsigögn alls efnis eru geymd í RecordSearch, en starfsmenn safnsins ákveða hvort lýsigögnin eru aðgengileg eða ekki. Efni getur þannig verið í takmarkaðri birtingu, eða ákvörðun um aðgengi ekki verið tekin, en samt er hægt að sjá lýsigögn þess í RecordSearch. Efni sem opnast, t.d. að loknum 30 árum, er fyrst fært á birtingarform og því næst flutt úr varðveislugeymslunni yfir í RecordSearch aðgangskerfið á enn einum hörðum disk (D5). Efni sem ekki er aðgengilegt á vef er afgreitt af starfsfólki gegn sérstökum beiðnum og er þá skoðað á lokuðum vélum innan safns.

NAA gerir ráð fyrir að varðveita um 5% af því stafræna efni sem verður til.³⁵ NAA tekur stefnumarkandi ákvarðanir um eyðingu efnis, en þegar ákvörðun hefur verið tekin getur skjalamyndari framfylgt henni. Enn sem komið er eru engar stofnanir farnar að skila efni

³³ National archives of Australia. *Digital preservation: digital preservation software applications*. Skoðað 29.10.2006. <http://www.naa.gov.au/recordkeeping/preservation/digital/applications.html>

³⁴ National archives of Australia. *XML data formats*. Skoðað 29.10.2006. http://www.naa.gov.au/recordkeeping/preservation/digital/xml_data_formats.html

³⁵ National archives of Australia. *Digital preservation: illuminating the past, guiding the future*. 2006. Bls. 7. http://www.naa.gov.au/recordkeeping/preservation/digital/XENA_brochure.pdf

reglulega, heldur hafa skil fyrst og fremst verið vegna neyðarráðstafana eða frá stofnunum sem lagðar hafa verið niður.

Eins og fram kemur hefur NAA skrifað eigin kerfi fyrir allt ferlið, sem byggð eru á opnum kerfum og stöðlum (öll kerfin eru opin og aðgengileg á netinu, nema RecordSearch):

- XENA (XML Electronic Normalising of Archives) færir efni yfir á varðveislusnið, og til baka yfir á birtingarsnið. XENA er forritað í Java.
- DPR (Digital Preservation Recorder) skráir lýsigögn um allt varðveisluferlið. Lýsigögn á hverju lokuðu svæði eru geymd í PostgreSQL gagnagrunni og færð yfir á diskum með efninu sem varðveita á. Lýsigögn DPR eru skilgreind af NAA, en hafa verið borin saman við staðla eins og PREMIS og eru talin ásættanlega skilgreind.
- Quest skráir lýsigögn um varðveitt efni í PostgreSQL gagnagrunni og er notað til að veita aðgang að efni, ýmist til að flytja yfir í RecordSearch eða til að birta á lokuðum vélum.
- RecordSearch geymir öll lýsigögn um allt efni, bæði stafrænt og annað efni. Einnig er efni geymt á stafrænu birtingarsniði í RecordSearch. Það á bæði við um stafrænt efni og annað efni sem gerð hafa verið stafræn afrit af, s.s. pappírsskjöl.

Öryggi í varðveisluferlinu er tryggt með því að nota lokuð net og takmarka aðgang að þeim.

4.2.2 National Library of Australia (NLA)

Þar sem ekki er um skilaskyldu að ræða á stafrænu efni semur NLA við einstaka útgefendur um skil til safnsins. Einnig varðveitir NLA vefefni, og þá aftur í samstarfi við einstaka útgefendur. NLA býr til eigið efni á TIFF sniði, en einnig Jpeg myndir fyrir birtingu á vef.

NLA hefur byggt talsverða innviði fyrir varðveislu stafræns efnis.³⁶ Þar sem ekki er um mjög viðkvæmt efni að ræða, er efnið aðeins geymt í einu kerfi sem aðgengilegt er á neti safnsins. Sérhæfð leitarvél er notuð til að veita aðgang út á við. Flestir þættir hugbúnaðarkerfisins, fyrir utan leitarvélinu, eru heimasníðaðir þar sem starfsfólk safnsins taldi engin kerfi henta þegar högunin var hönnuð. Almenn má segja að högunin sé í nokkurri þróun; verið er að endurbæta sum kerfi og skoða hvort þurfi að skipta út öðrum.

Efni er varðveitt í DOSS (Digital Object Storage System), sem er í aðalatriðum skráarkerfi á SAM-FS kerfi frá Sun og sameinar diskstöð og bandstöð. Diskstöðin gefur góð afköst á því efni sem mest er sótt, t.d. Jpeg birtingareintök, en bandstöðin gefur möguleika á að geyma mikið gagnamagn, t.d. TIFF varðveislueintökin. Til að bæta líkur á langtímarvarðveislu eru geymd nokkur auka-eintök (e. redundant copies) af öllu efni. Alls er stafræna safnið um 25TB, en um þrefalt meira efni er geymt á segulböndum vegna auka-eintaka. Afrit eru geymd utan safnsins.

DCM (Digital Collections Management) er annað heimasníðað kerfi, sem sér um vinnuferli við skráningu efnis og yfirfærslu eldra efnis á stafrænt snið. DCM geymir lýsigögn í Oracle gagnagrunni, en efnið sjálft í DOSS. Verið er að skoða notkun Fedora kerfisins til að vista saman lýsigögn og efni, þannig að DCM kerfið geti einbeitt sér að verkferlinu í stað skipulagningar á undirliggjandi geymslu. Hvorki DCM né DOSS eru opin kerfi. Starfsfólk safnsins hefur einnig skoðað DSpace að einhverju marki, en til stendur að prófa Fedora á næstu mánuðum (sjá kafla 4.10).

³⁶ National Library of Australia. *NLA digital library: logical systems architecture*. Skoðað 29.10.2006. <http://www.nla.gov.au/dsp/architecture.jpg>

Aðgengi að efni er gegnum TeraText, sem er textaleitarvél sem er notuð til að leita að efni úr DCM/DOSS og öðrum kerfum, sjá t.d. PictureAustralia.³⁷ Þessi leitarvél er framleidd af InQuirion, áströlsku fyrirtæki sem leitt er af Ron Sacks-Davis, frægum rannsakanda á sviði textaleitar. Samkvæmt upplýsingum frá safninu er verið að skoða aðrar lausnir fyrir leitina.

Vefefni er geymt í DOSS en stjórnað með PANDAS kerfinu sem þróað var í PANDORA verkefninu (Preserving and Accessing Networked Documentary Resources of Australia).³⁸ Vefefni er geymt annars vegar í upprunalega horfi, en hins vegar eru geymd eintök sem eru löguð til birtingar. Ekki hafa verið gerðar neinar ráðstafanir til að breyta sniði vefefnis sem úreldist.

NLA hefur haldið úti PADI vefsíðunni (Preserving Access to Digital Information)³⁹, sem hefur að geyma yfirlit yfir margvíslegt efni sem tengist langtímarvarðveislu stafrænna gagna.

4.3 Bretland

Í Bretlandi er varðveisluhlutverk bóka- og skjalasafna nokkuð dreift. The British Library (BL) og The National Archives (NA) gegna lykilhlutverki, en fjölmörg héraðs- og háskólasöfn skipta líka máli. Tveimur stórum samvinnuverkefnum er ætlað að sameina krafta þessarar safna, JISC (Joint Information Systems Committee),⁴⁰ og DPC (Digital Preservation Coalition).⁴¹ DPC gaf út á árinu skýrslu sem veitir gott yfirlit yfir stöðu verkefna í Bretlandi.⁴²

The National Archives er þó í forystuhlutverki í Bretlandi og hlaut viðurkenningu árið 2004 fyrir starfsemi sína á sviði varðveislu stafrænna gagna. Í ljósi þess, og vegna þess að lögð var áhersla á skjalasöfn sem varðveita stjórnsýslugögn, verður hér fjallað nánar um The National Archives en ekki The British Library.

Vegna nýlegra lagabreytinga er í undirbúningi að taka við opinberu efni sem fyrst eftir að það verður óvirkt. Þessar lagabreytingar gera það að verkum að í stað þess að efni verði aðgengilegt 30 árum eftir að það varð til, er það aðgengilegt strax. Á þessu eru þó takmarkanir, eins og t.d. með efni sem geymir persónulegar upplýsingar eða upplýsingar er varða þjóðaröryggi. Meira um þessi lög er hægt að sjá á Internetinu.⁴³

4.3.1 The National Archives (NA)

Fram að þessu hefur The National Archives einkum verið að taka við sérhæfðu, stafrænu efni sem hefur þótt sérlega áhugavert, en um leið líklegt til að úreldast. Sem dæmi má nefna vefi um athuganir sérstakra nefnda stjórnsýslunnar og hefur þetta efni verið á ýmsu sniði. Til dæmis um þetta er efni frá rannsókn á skipsskaða frá 1980,⁴⁴ þar sem varðveita þurfti rúmlega 2.000 klukkustundir af neðansjávarmyndum, næstum 140.000 ljósmyndir, sönargögn og sýndarveruleikalíkan.

³⁷ Picture Australia. <http://www.pictureaustralia.org/>

³⁸ PANDORA. <http://pandora.nla.gov.au/index.html>

³⁹ National library of Australia. *Preserving Access to Digital Information*. <http://www.nla.gov.au/padi/>

⁴⁰ Joint information systems committee. <http://www.jisc.ac.uk/>

⁴¹ Digital preservation coalition. <http://www.dpconline.org/>

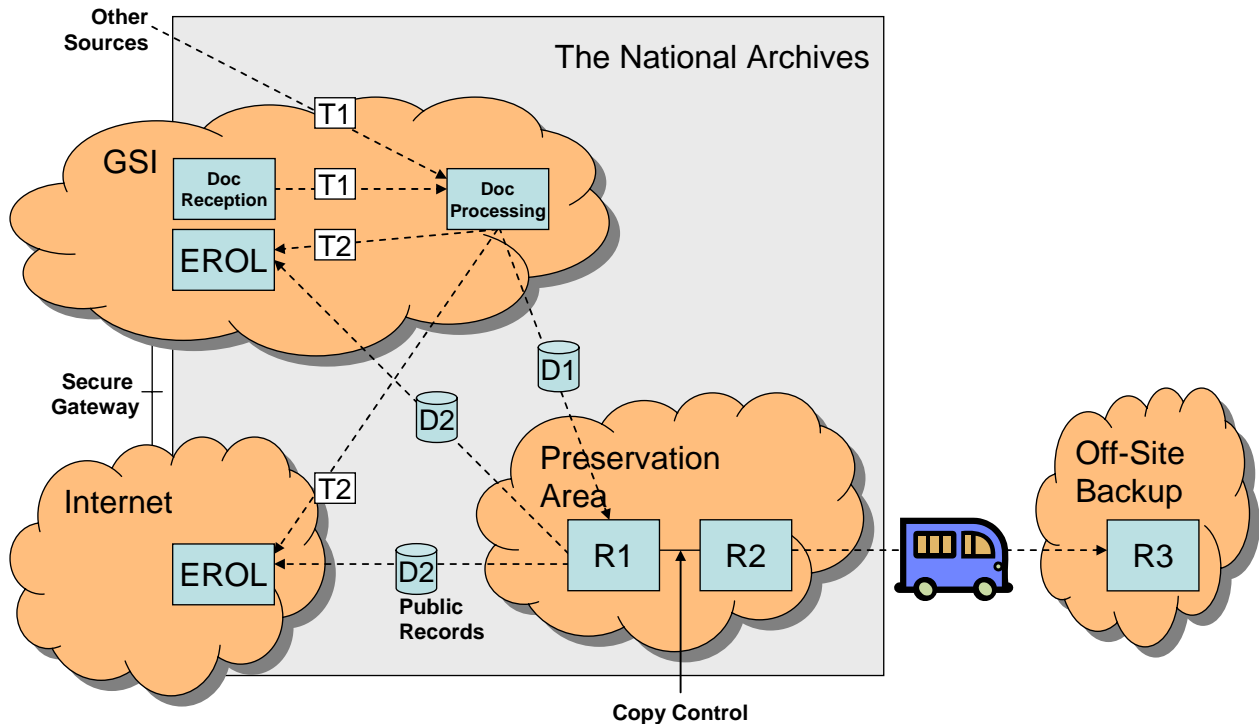
⁴² Waller, M. og Sharpe, R. *Mind the gap*. 2006. <http://www.dpconline.org/docs/reports/uknamindthegap.pdf>

⁴³ Department for constitutional affairs. *Freedom of information*. Skoðað 29.10.2006.

<http://www.dca.gov.uk/foi/index.htm>

⁴⁴ Brown, A. *Preserving the digital heritage*. 2003. <http://www.nationalarchives.gov.uk/documents/brown.pdf>

Verið er að skipuleggja heildarferli til að tryggja öruggt flæði heildstæðs efnis inn í safnið.⁴⁵ Gert ráð fyrir að flokka efni á skipulegan hátt hjá stofnunum til að minnka efnismagnið og velja það sem mestu máli skiptir. Lýsigögn sem fylgja efninu innihalda sammengi ýmissa lýsigagnastaðla, þ.m.t. PREMIS.



Mynd 4-2 : Yfirlit yfir varðveisluferli hjá The National Archives í Bretlandi

Mynd 4-2 gefur yfirlit yfir varðveisluferli safnsins, sem er í aðalatriðum tvískipt.

Annars vegar er um að ræða ferli fyrir inntöku efnis. Efni er tekið inn í gegnum lokað innra net stjórnsýslunnar (GSI, Government Secure Internet, sjá kafla 2.3.2) eða á geymslumiðlum (örvar merktar T1 á Mynd 4-2). Efnið er vírusleitað, tékksummur athugaðar og farið almennt yfir heilleika efnisins; svo er það sett í sóttkví í fjórar vikur og athugað aftur. Að þessu loknu eru birtingarútgáfur opins efnis settar beint inn í kerfið EROL, sem lýst er hér að neðan (T2). Loks er efnið fært til varðveisluna á lausum diskum (D1).

Hins vegar er um að ræða varðveisluna sjálfa, sem fer fram á neti sem er ótengt umheiminum. Tvö varðveislukerfi eru í notkun, bæði byggð á sama vél- og hugbúnaði. Stafræna efnið er fyrst sett inn í annað kerfið (R1) og svo afritað yfir á hitt (R2). Þegar beiðni kemur um aðgang að lokuðu efni er tekin ákvörðun um birtingu. Ef opna á efnið, þá er það fært yfir í aðgangskerfið EROL (D2). Ef efnið á aðeins að birta viðkomandi einstaklingi, er það afritað eða prentað, og afhent honum.

⁴⁵ The national archives. *Seamless flow*. Skoðað 29.10.2006.
http://www.nationalarchives.gov.uk/electronicrecords/seamless_flow/default.htm

The National Archives lét smíða fyrir sig kerfið Digital Archiving System sem notað er fyrir inntöku og varðveislu efnis (Tessella Support Services plc).⁴⁶ Þetta kerfi hefur eftirfarandi eiginleika:

- Java applet er notað til að skrá efni og lýsigögn.
- XML vefviðmót er notað til að hlaða inn efni, stutt af Oracle Application Server.
- Oracle gagnagrunnur er notaður til að geyma lýsigögn.
- Compaq RAID server er notaður til að geyma efnið þegar það er skoðað í upphafi.
- Efnið er geymt á böndum og/eða diskum með FileTek StorHouse/RFS.
- Undirliggjandi er StorageTek L20 Tape Library.

Kerfið hlaut alþjóðlegu verðlaunin Pilgrim Trust Perservation Award árið 2004. Kerfið sem þjóðskjalasafnið í Bandaríkjunum hefur í smíðum mun einnig byggja á FileTek StorHouse/RFS.

FileTek StorHouse/RFS er í raun skráarkerfi sem er byggt ofan á venslaðan gagnagrunn (RFS= Relational File System, svipað Oracle Internet File System, iFS). Það er velþekkt vandamál við skráarkerfi sem byggð eru á hefðbundinn hátt að þau ráða ekki vel við mikinn fjölda skjala. Með því að geyma upplýsingar um skjölin í gagnagrunni og nota hann til að finna skjölin og skila þeim til baka, er hægt að nota leitartækni venslaða grunnsins til að auka skalanleikann.

Í fréttatilkynningu frá FileTek kemur fram að kerfið hafi verið notað til að geyma milljarð skjala (e. records) á einu skráarsvæði.⁴⁷ StorHouse/RFS er skrifað til að vera óháð geymslumiðli og getur því notað segulbandsstöðvar eða diskastæður eftir aðstæðum. Kerfið er sagt bjóða upp á auðveldla færslu gagna milli miðla eftir því sem geymsluformið þróast, sem er lykiltríði við langtímarvarðveislu.

FileTek StorHouse/RFS kerfið er m.a. í notkun hjá The National Archives í Bretlandi, og Veðurstofu Bretlands, The Met Office⁴⁸, til að hýsa stór skjalasöfn. Í öllum tilvikum er StorHouse/RFS notað sem stór skjalageymsla, þar sem kerfi er smíðað ofan á sem sér um lýsigögn og leit.

Eins og áður var nefnt, fer aðgangur fram gegnum kerfi sem heitir EROL (Electronic Records On-Line). Opið efni og lýsigögn eru geymd á EROL þjóni, sem er aðgengilegur á Internetinu og er notaður við vefleit í opnu efni safnsins. Efni er sett beint í EROL við inntöku ef það á við, eða fært seinna yfir í EROL á segulmiðlum eftir ákvörðunum starfsfólks safnsins. Opið efni er almennt aðgengilegt á vefnum, en efni á sérstökum sniðum og mjög umfangsmikið efni þarf að panta sérstaklega og er þá afhent á geisladiskum eða segulmiðlum eftir atvikum.

Verið er að vinna að nýrri útgáfu EROL sem verður aðgengileg á lokuðu innra neti stjórnsýslunnar (GSI, Government Secure Internet⁴⁹) og mun innihalda allt efni. Aðgangsstýringar verða notaðar, þannig að hver starfmaður stjórnsýslunnar sjái aðeins það efni sem honum/henni er ætlað. Þetta er eina tilvikið sem rekist var á, þar sem lokað efni er aðgengilegt utan frá.

⁴⁶ Tessella support services. <http://www.tessella.com/>

⁴⁷ FileTek. *FileTek Successfully Writes One Billion Files to a Single StorHouse/RFS File System*. 2005. Skoðað 29.01.2007.

⁴⁸ <http://www.met-office.gov.uk/>

⁴⁹ http://www.cw.com/docs/services/product_pdfs/gsc-ss.pdf

4.4 Bandaríkin

Í Bandaríkjunum er varsla stafræns efnis (og annars efnis) óhemju dreifð og fjölmargir aðilar koma að henni af hálfu ríkisstjórnar, einstakra fylkja og smærri stjórnsýslueininga. Þó er Library of Congress (LoC) í fararbroddi fyrir bókasöfn og National Archives and Records Administration (NARA) fyrir skjalasöfn.

LoC hefur rekið verkefnið NDIIPP⁵⁰ (National Digital Information Infrastructure and Preservation Program). Þrátt fyrir að fjölmargir aðilar hafi komið að því verkefni um langa hríð verður ekki séð að mikið haldbært hafi komið út úr því enn sem komið er.⁵¹ Því verður hér einkum fjallað um NARA.

4.4.1 National Archives and Records Administration (NARA)

NARA hefur skilgreint verkefni, kallað Electronic Records Archives (ERA), sem snýst um að búa til kerfi til að varðveita og veita aðgang að stafrænu efni.⁵²

NARA hefur staðið fyrir fjölmörgum rannsóknarverkefnum í samstarfi við háskóla og stofnanir. Sem dæmi má nefna að í Bandaríkjunum er mikill áhugi á að varðveita tilraunagögn sem orðið hafa til fyrir styrki frá National Science Foundation, og hafa verið unnin verkefni í samstarfi við San Diego Supercomputer Center og háskóla, um notkun „grid“ aðferðafræði við aðgang að slíku efni, sjá t.d. grein um aðferðafræðina eftir JaJa et al. (2004).⁵³ Grid aðferðafræðin snýst í stuttu máli um að veita notendum kerfa aðgang að gagnaveitum og reikniveitum hvar sem þær kann að vera að finna, án þess að notandinn þurfi að vita hvað er að gerast undir yfirborðinu. Þetta efni er hins vegar mjög opið og áfram geymt hjá viðkomandi stofnunum.

Til að geyma viðkvæmara stafrænt efni sem verður til hjá ríkisstofnunum og hýst er hjá NARA, hefur stofnunin boðið út smíði varðveislukerfis.⁵⁴ Lausnin, sem vinna hefur hafist við, er unnin af Lockheed Martin og mun nota sömu undirliggjandi tækni og The National Archives í Bretlandi notar, nefnilega FileTek StorHouse/RFS. Gert er ráð fyrir að það kerfi verði tekið í notkun árið 2007. Verðgildi samningsins er rúmlega 20 milljarðar íslenskra króna yfir sex ára tímabil.

Ekki barst svar við frekari spurningum sem sendar voru til NARA, en ljóst að áhugavert verður að fylgjast með framgangi mála hjá þeim.

4.5 Danmörk

Í Danmörku er varðveisluhlutverk bóka- og skjalasafna nokkuð hefðbundið. Nokkur bókasöfn sjá saman um varðveislu útgefins efnis, og hafa samvinnu á þann hátt að þau geyma sama efni, en á mismunandi vél- og hugbúnaði, til að tryggja sig gegn úreldingu. Ekki verður fjallað nánar um bókasöfnin hér, eins og áður hefur komið fram.

Þjóðskjalasafn Danmerkur, Statens Arkiver (SA), hefur tekið við stafrænu efni í langan tíma og mótað ákveðna stefnu í varðveislumálum sem er að vissu marki ólík öðrum löndum sem skoðuð voru. Hér á eftir verður farið nánar yfir þessa stefnu.

⁵⁰ NDIIPP. <http://www.digitalpreservation.gov/>

⁵¹ Gladney, H.M. *Digital preservation*. 2005. http://home.pacbell.net/hgladney/ddq_5_2.htm

⁵² The national archives. *ERA Documentation*. Skoðað 13.11.2006.

<http://www.archives.gov/era/about/documentation.html>

⁵³ JaJa, J. *Digital archiving and long term preservation*. 2004. Skoðað 13.11.2006.

<http://www.archives.gov/era/papers/thic-04.html>

⁵⁴ The national archives. *National archives names Lockheed Martin to build archives of the future*. 2005.

<http://www.archives.gov/press/press-releases/2005/nr05-112.html>

4.5.1 Statens Arkiver

Þegar ferli stafræns efnis hjá SA er skoðað „af löngu færi“ er það mjög líkt og hjá öðrum söfnum. Efni er flutt frá skjalamyndurum á segulmiðlum, yfirfarið á lokuðu kerfi, geymt um tíma í einangrun, og svo yfirfarið aftur. Þá er efnið geymt í nokkrum eintökum á fleiri en einni tegund miðils. Með reglulegu, og skipulögðu, millibili er efnið fært milli miðla, áður en mikil hætta er á að miðlarnir skemmist eða úreldist. Ekki hefur verið mikið um aðgang að efni⁵⁵ og ekki hefur verið mótað ferli þar um, en verkefni er í gangi sem á að ljúka 2008.⁵⁶

Þegar heildarferlið er skoðað nánar koma þó í ljós nokkur atriði sem eru athyglisverð.

Lagarammi SA er sterkur og safnið getur gert talsverðar kröfur til efnisins við afhendingu. Safnið hefur sett ákveðnar reglur um efnissnið sem allir skjalamyndarar verða að uppfylla. Sem dæmi má nefna að textaskjölum og myndum er skilað á TIFF sniði, með texta á textasniði þar sem við á. Safnið hefur hingað til ekki samþykkt PDF sem varðveislumiðil, og telur PDF/A enn vanta útbreiðslu til að teljast ásættanlegur staðall, einkum hvað varðar hugbúnað sem styður sniðið.

Hvað varðar varðveislumiðla, þá er SA eina safnið sem skýrsluhöfundar fundu sem geymir varðveisluefni á öðrum miðlum en segulmiðlum. Allt efni SA er varðveitt á tveimur eintökum af geisladiskum (CD-R) frá mismunandi framleiðendum, en nýlega var einnig bætt við diskastæðu sem þriðja varðveislumiðlinum.

SA hefur komið sér upp umtalsverðum búnaði til að meðhöndla geisladiskana. Þannig á safnið tvö mismunandi tæki til að prófa diskagerðir, afkastamikil tæki til að skrifa diska, og tæki til að fylgjast skipulega með diskum og hvernig þeir eldast. Auk þess hefur safnið skrifað hugbúnað fyrir þessi tæki, sem þeir telja vera mun betri en hugbúnað framleiðenda í mörgum tilfellum.

Núna á safnið um 6.400 geisladiska (3.200 í tveimur eintökum hvern) og hefur nýlega verið farið í gegnum allt safnið og það flutt yfir á nýja diska. Allt efni er nú einnig til á RAID 6 diskastæðu (í RAID 6 diskastæðu eru tveir aukadiskar til að geyma upplýsingar til villuprófana). Á næsta ári er gert ráð fyrir að taka á móti mun meira gagnamagni en áður og er safnið því að færa sig yfir í DVD-R diska og LTO-3 segulbönd sem varðveislumiðla.

Varðveisla SA á geisladiskum hefur haft umtalsverð áhrif á snið efnis sem skilað er til safnsins. Gerð er krafa um að efnið sé bútað niður í hluta sem hver um sig passar á einn geisladisk.⁵⁷ ⁵⁸ Áður var efni skilað á geisladiskum, en safnið hefur nýlega breytt þeirri reglu og fer núna fram á skil á segulmiðlum til að auka afköst við úrvinnsluna. SA hefur skrifað hugbúnað til að fara yfir afhent efni og greina hvort það sé í samræmi við kröfur. Ef svo er ekki, þarf skjalamyndari að skila aftur. Safnið hefur unnið með skjalamyndurum til að kenna þeim að skila á réttan hátt.

Loks hefur SA lagt ríka áherslu á skil á gagnagrunnum, ríkari áherslu en annars staðar hefur sést. Gagnagrunnum er varpað í flatar skrár á einföldu sniði („fyrsta normalformi“), og skráðar upplýsingar um atriði eins og vensl milli taflna, helstu SQL fyrirspurnir sem notaðar eru til að ná upplýsingum úr grunninum o.s.frv. Þessar upplýsingar eiga að vera nægar til að

⁵⁵ Madsen, K. *Debatoplæg om benyttelse af de elektroniske arkivalier*. 2006.

http://www.skjalasafn.is/Syningar/img/Debatoplaegisland_KM.ppt

⁵⁶ Jespersen, L.S. *Tilgængeliggørelse af elektroniske arkivalier*. 2006.

http://www.skjalasafn.is/Syningar/img/Projektpraesentation_LSJ.ppt

⁵⁷ Edelholt, P. 2006. http://www.skjalasafn.is/Syningar/img/Peter_Edelholt.ppt

⁵⁸ Olsen, R.M. *Filformater*. 2000.

http://kum.inforce.dk/graphics/kum/billeder/temaer/tema_bevaring/bilag_til_bevaring_af_kulturarven/bilag_13.doc

reyndur tölvunarfræðingur geti lesið töflurnar inn í venslaðan gagnagrunn og unnið með þær. Í reglum safnsins (reglur Þjóðskjalasafns Íslands eru nokkuð bein þýðing á þeim) er skýrt tekið fram hvernig snið grunnanna skal vera og SA hefur skrifað búnað til að prófa grunnana, auk hugbúnaðar til að skoða einstakar töflur og færslur.

Lýsigögn voru skilgreind af SA áður en helsta alþjóðlega vinnan hófst við stöðlun lýsigagna og því fer safnið ekki eftir lýsistöðlum. Þeir telja að staðlarnir geri ráð fyrir að stafrænt efni sé betur skipulagt en það er í raun og að nánast megi fara með stafrænt efni eins og bækur. Ljóst er að athugasemd þeirra á sérstaklega við um gagnagrunnana.

Aðgengi að efninu hefur verið takmarkað, bæði hvað varðar beiðni um aðgang og hvað varðar aðstöðuna. Leit fer fram með aðstoð starfsfólks safnsins og eftir að efni er fundið er ekki auðvelt að skoða það. Starfsfólk safnsins viðurkennir að núna þurfi nánast tölvunarfræðigráðu til að geta skoðað efnið.⁵⁹ Hins vegar hefur safnið sett í gang verkefni sem keyrt verður frá 2006-2008 sem hefur það að markmiði að gera efnið aðgengilegra.

4.6 Svíþjóð

Í Svíþjóð er staðan svipuð og víða annars staðar. Riksarkivet tekur við stafrænu efni, auk þess sem dreift kerfi bóka- og skjalasafna þarf að huga að slíku efni. Þá hefur verið stofnaður samstarfsvettvangur stofnana og háskóla, Projektet Långsiktigt Digitalt Bevarande (LDB), sem sameinar krafta við rannsóknir og þróun langtímarvarðveislu. Um þetta verkefni verður rætt hér á eftir.

4.6.1 Riksarkivet

Riksarkivet hefur tekið við stafrænum efni frá 1970, og á árinu 2005 tók safnið við um 25TB. Fyrir tveimur árum fjárfesti safnið í geymslukerfi sem sameinar diska og bönd á líkan hátt og að ofan er lýst (t.d. fyrir Bretland).

Ekki hefur verið virk langtímarvarsla, en Riksarkivet tekur þátt í LDB-verkefninu, þar sem umtalsverð markmiðasetning hefur farið fram. PDF/A staðallinn hefur verið samþykktur af safninu, lýsigögn skilgreind og eru þau blanda af ýmsum stöðlum.

Í áhugaverðri grein eftir Jonas Palm⁶⁰ lýsir hann kostnaðargreiningu á varðveislu stafræns efnis, sem sýnir vel hversu dýr slík varðveisla er. Þótt diskarými verði sífellt ódýrara, þá felist mesti kostnaðurinn í öðrum búnaði og ekki síður umsýslu hvers konar, og sá kostnaður er meiri og fer vaxandi. Í lok greinarinnar nefnir hann að athuga þurfi hvort til greina komi að nota tölvuskrafaðar örfilmur (COM, Computer Output Microfilm, sjá umræðu í kafla 3.4) sem varðveislumiðil, þar sem þær endist lengi og auðvelt sé að skanna þær inn ef þarf.

4.6.2 LDB-verkefnið

LDB-verkefnið er samstarfsverkefni Riksarkivet, Kungliga Biblioteket, Luleå Tekniska Universitet, og fleiri aðila.⁶¹ Verkefnið hefur þrjú meginmarkmið.

Fyrsta markmið verkefnisins er að stunda rannsóknir á sviði langtímarvarðveislu stafræns efnis. Þannig hafa verið teknar saman og gefnar út nokkrar samantektir um stöðu mála, auk þess sem nemendaverkefni hafa verið unnin, m.a. um varðveislu skjámynda. Ein afleiðing af þessu starfi er að til verður hópur fólks með menntun á sviði langtímarvarðveislu, sem getur

⁵⁹ Nielsen, A.B. *Preservation of Electronic Records*. 2004. Bls. 7.

http://www.wien2004.ica.org/imagesUpload/pres_190_NIELSEN_A_DEN01E.pdf

⁶⁰ Palm, J. *The digital black hole*. 2006? http://www.tape-online.net/docs/palm_black_hole.pdf

⁶¹ <http://ldb.project.ltu.se/>

ráðist til starfa á sviðinu í Svíþjóð. Til lengri tíma litið gæti þetta verið mikilvægasta afurð verkefnisins.

Annað markmið verkefnisins er að leggja drög að varðveislukerfi. Í póstsamskiptum kom fram að verið er að leggja lokahönd á uppdrátt að slíku kerfi í samstarfi við Riksarkivet, og gert ráð fyrir að tilraunaútgáfa verði tilbúin í árslok 2007. Hugmyndin er að kaupa kerfi síðar, því LDB-verkefninu er ekki ætlað að sinna hugbúnaðarþróun.

Þriðja og síðasta markmiðið er að efla samstarf stofnana og skóla innan Svíþjóðar. Í því skyni eru m.a. haldin námskeið.

4.7 Holland

Í Hollandi er varðveislulutverk bóka- og skjalasafna nokkuð dreift. Koninklijke Bibliotheek (KB) og Nationaal Archief (NA) gegna þó lykilhlutverki, og einkum hefur KB verið í fararbroddi í útfærslu og rannsóknum á langtíma varðveislu gagna.

4.7.1 Koninklijke Bibliotheek (KB)

Ekki er um skilaskyldu til bókasafna að ræða í Hollandi. Hins vegar eru margir stórir útgefendur vísindatímarita með aðsetur þar, og safnið hefur safnað þessum vísindatímaritum gegnum tíðina. Þegar þessir útgefendur tóku upp rafræna útgáfu og dreifingu vísindatímarita, hætti safnið að fá prentuð eintök til varðveislu og fór því að huga að stafrænni varðveislu. Safnið hefur gert samninga við nokkra útgefendur með tvennt í huga: Annars vegar að veita aðgang innan safns að þessu efni til frambúðar, og hins vegar að geta tryggt varðveislu ritanna ef útgefandi verður fyrir skakkaföllum og tynir stafrænu efni sínu.

Aðferðafræði KB við langtíma varðveislu er þrískipt:⁶²

- Í fyrsta lagi að búa til örugga geymslu, sem öllu stafrænu efni safnsins er safnað í.
- Í öðru lagi að tryggja örugga endurnýjun bitanna sem geymdir eru, með því að tryggja uppfærslu kerfa við breytingar á geymslumiðlum.
- Í þriðja lagi að tryggja varðveislu innihalds efnisins, með því að tryggja að ávallt sé hægt að birta efnið.

Segja má að fyrstu tveimur markmiðunum hafi verið náð og verið sé að vinna að því þriðja.

Eftir talsverða undirbúningsvinnu gerði KB samning við IBM árið 2000, um þróun og smíði á öruggri geymslu, e-Depot. Markmiðið var að nota tilbúna hugbúnaðarpakka að eins miklu marki og hægt væri, en nýskrifa það sem vantaði upp á. Samningurinn var þannig samansettur að IBM héldi öllum réttindum á hugbúnaðinum og gæti þannig markaðssett hann sem nýja vöru að vinnu lokinni. Þannig væru hagsmunir IBM við verkið mun stærri. Úr þessu varð kerfi sem er kallað Digital Information Archiving System eða DIAS, en þetta kerfi hefur verið í notkun hjá KB síðan 2002.

DIAS kerfið er skrifað í samræmi við OAIS staðalinn og notar marga tilbúna pakka frá IBM, eins og Content Manager, Tivoli Storage Manager, WebSphere Application Server, DB2 gagnasafnskerfið, og Business Objects. Þættir sem lutu að inntöku efnis og birtingu efnis voru nýskrifaðir. Kerfið keyrir á tveimur IBM RS/6000 vélum, en undirliggjandi er SAN stæða byggð á diskastæðu, bandstöð, og „optical library“ stöð. DIAS kerfið útfærir fyrstu tvo þættina að ofan, örugga geymslu og örugga endurnýjun bita.

⁶² Steenbakkens, J.F. *Digital archiving in the twenty-first century*. 2005. Bls. 42.

DIAS kerfið (eða e-Depot) geymir stafræna efnið en lýsigögn eru geymd í eldra bókasafnskerfi KB, sem endurbætt var til þess að leita í lýsigögnunum og ná í efni í DIAS. Ástæðan fyrir aðskilnaði efnis og lýsigagna er sögð sú að þannig er hægt að breyta lýsigögnum án þess að snerta efnið sjálft, sem minnkar líkur á að bitar brenglist í meðförum. Að auki eru nefndar tvær hagnýtar ástæður. Í fyrsta lagi er mikil þróun á lýsigögnum og þeim hefur stundum öllum verið breytt á einu bretti, t.d. til að setja inn DOI (Digital Object Identifier) merkingu. Í öðru lagi er efnið sem geymt er mjög umfangsmikið, þannig að ef efni og lýsigögn væru geymd saman þyrfti að fara gegnum mjög mikið gagnamagn til að gera slíkar breytingar, sem gæti tekið verulegan tíma og því meiri sem meira hefur safnast upp af efni.

Efni er tekið inn og varðveitt á því sniði sem það er gefið út á. Hins vegar stendur til, innan árs, að búa til eintök af öllu innteknu efni á traustum varðveislusniðum (e. normalization), þ.e. setja efni á stöðluð og samræmd snið eins og PDF, TIFF eða Jpeg. Upprunalegu eintökin verða varðveitt áfram. Ekki eru geymd umfram afrit (e. redundant copies) af efni, en afritataka er notuð.

Eitt eintak af DIAS er í gangi, með hefðbundinni afritatöku. Hins vegar er ráðgert að setja upp afrit utan safnsins. KB notar útgáfu 1.3 af DIAS en gerir ráð fyrir að fara í útgáfu 2.0 í upphafi næsta árs. Þess má geta að nokkur söfn í Þýskalandi eru að taka DIAS 2.0 í notkun (KOPAL verkefnið⁶³).

Til að tryggja að ávallt verði hægt að birta efni er KB að þróa verulega áhugaverða aðferðafræði. Þessi aðferðafræði felst í tveimur lykilþáttum:⁶⁴

- Að tryggja að ávallt sé leið frá sniðinu yfir í birtingu (e. view path), þ.e. að ávallt sé til hugbúnaður og vélbúnaður sem saman getur birt efnið (t.d. Adobe Acrobat Reader/Windows 2000/PC). Markmiðið er raunar að ávallt séu til a.m.k. tvær leiðir fyrir hvert skjalasnið. Þetta er gert til að önnur sé eftir ef hin úreldist, þannig að hægt er að skrifa nýja leið og bera saman við þá sem fyrir er.
- Að fylgjast með þessum leiðum, hvenær þær úreldast og hvernig á að bregðast við því. T.d. ef fyrirtæki sem heldur við birtingarhugbúnaði verður gjaldþrota, þá þarf að athuga hvort þessi hugbúnaður gegni lykilhlutverki í annarri af tveimur leiðum fyrir eitthvert snið. Ef svo er, þá þarf að finna aðra leið fyrir það snið.

Verið er að skrifa einingu, Preservation Manager, í DIAS sem á að stýra þessum birtingarleiðum og gefa viðvaranir ef einhverjar þeirra eru í hættu og þar með efni á viðkomandi sniði. Gert er ráð fyrir að sú eining verði komin í notkun innan árs.

Birtingarleiðirnar sjálfar geta verið mismunandi eftir mismunandi efni. Verið er að rannsaka bæði yfirlæslu (e. migration) og hermun (e. emulation) innan KB, og gert ráð fyrir að sambland þessarar aðferða verði nauðsynlegt til að ná árangri til langs tíma.

Fyrir óbreytanlegt efni, t.d. textaskjöl, getur komið til greina að færa efnið á nýtt snið sem líklegra er til varðveislu. Dæmi um þetta er sú aðferð sem að ofan er nefnd, að koma öllu efni á traust varðveislusnið við inntöku.

Efni sem hefur „hegðun“ (t.d. leikir og önnur forrit) er meðhöndlað þannig núna, að það er fyrst sett upp á viðmiðunarvél (e. reference work-station, RWS) og því næst er tekið afrit af hörðum diskum viðmiðunarvélarinnar í heild og geymt í DIAS. Þessi viðmiðunarvél er þá dæmi

⁶³ <http://kopal.langzeitarchivierung.de/index.php/en>

⁶⁴ Oltmans, E., van Diessen, R. og van Wijngaarden, H. *Preservation functionality in a digital archive*. 2004.

um birtingarleið, sem halda má við eins og öðrum birtingarleiðum. Markmið KB, til lengri tíma, er að geta notað hermun til að halda slíkum hugbúnaði við.

KB er að hefja verkefni um vefvarðveislu⁶⁵ og mun geyma sóttar vefsíður í DIAS kerfinu.

4.7.2 Nationaal Archief

National Archief (NA) hefur hafið samstarf við KB um langtímarvarðveislu og mun safnið taka DIAS í notkun innan skamms. Safnið hefur einnig staðið fyrir verkefninu Digitale Duurzaamheit („Stafrænt langlífi“), en það hefur snúist um að skoða gagnasnið og fleiri þætti stafrænnar varðveislu. Til dæmis hafa þar verið skilgreind varðveislusnið fyrir flókið efni eins og skjöl úr töflureiknum.

Safnið hefur borið saman þrjú kerfi til stafrænnar langtímarvarðveislu:

- eDepot, sem byggt er á Dspace (sjá kafla 4.10) og smíðað af skjalasafni í Rotterdam.
- Digital Archiving System kerfið, sem smíðað var fyrir National Archives í Bretlandi.
- DIAS, sem notað hefur verið við Koninklijke Bibliotheek í Hollandi.

Niðurstaða safnsins er að skrifa þurfi nýtt kerfi, sem líkist þó Digital Archiving System mjög. Safnið mun hafa kerfið á lokuðu neti.

4.8 Kanada

Library and Archives Canada (LAC) er ein stofnun síðan 2004, sem sér um varðveislu fyrir allar tegundir skjala.

4.8.1 Library and Archives Canada

Af vef LAC að dæma er stofnunin ekki komin langt á leið í varðveislu stafræns efnis. Rætt er um að þróa varðveislukerfið EAPS (Electronic Archives Preservation System) og að ná færni í meðhöndlun stafrænna gagna á árunum 2005-2010. Einnig er rætt um nauðsyn þess að sameina söfn eldri stofnana í eitt safn (e. one collection).

Mikil vinna fer fram þessa dagana í Kanada við að móta innviði fyrir stafræna varðveislu (e. digital preservation infrastructure).⁶⁶ Sú vinna er mjög áhugaverð, og sambærileg vinna þyrfti að fara fram á Íslandi. Þegar leitað er eftir þeim fordæmum sem lengst eru á veg komin, þá er LAC hins vegar greinilega ekki meðal þeirra.

4.9 Evrópusambandið

Evrópusambandið hefur mikinn áhuga á eflingu stafrænnar langtímarvarðveislu innan ríkja sambandsins og á eigin vegum. Sem hluti af skammtímastefnu sambandsins hefur verið mótuð stefnan „i2010: Digital Libraries Initiative“⁶⁷, sem snýst um varðveislu menningar- og vísindaarfs sambandsríkjanna. Önnur skyld stefna til 2010 snýst um eflingu rafrænnar stjórnsýslu. Til að efla stafræna langtímarvarðveislu, stendur sambandið fyrir ýmsum verkefnum, og verða nokkur þeirra nefnd hér.

Fyrst má nefna að í þeim þætti 7. rammaáætlunar sem lýtur að upplýsingatækni, snýst ein af sjö höfuðáherslum („4. challenge“) um stafræn söfn og innihald þeirra, þar með talið stafræna

⁶⁵ Koninklijke Bibliotheek. *Web archiving*. Skoðað 29.10.2006.

http://www.kb.nl/hrd/dd/dd_projecten/projecten_webarchivering-en.html

⁶⁶ LAC. *Toward a Canadian digital information strategy*. Skoðað 29.10.2006.

<http://www.collectionscanada.ca/scin/index-e.html>

⁶⁷ http://ec.europa.eu/information_society/activities/digital_libraries/index_en.htm

langtímavarðveislu.⁶⁸ Næsta kall eftir verkefnum undir þessari áherslu er 2. kall, sem opnað verður í maí/júní 2007 og lokað í september/október 2007.

Í 6. rammaáætlun, 5. kalli, var einnig lögð áhersla á stafræna langtímavarðveislu. Tvö stór verkefni voru styrkt, CASPAR og PLANETS. CASPAR verkefnið snýst um varðveislu menningar-, lista-, og vísindaefnis og snýr meira að aðgengi og framsetningu efnis.⁶⁹ PLANETS verkefnið snýr meira að smíði verkfæra og aðferða fyrir langtímavarðveisluna sjálfa.⁷⁰ Bæði verkefni eru mjög áhugaverð, en þó einkum PLANETS verkefnið sem hefur mjög sterka aðila innan sinna vébanda.⁷¹ Að auki var styrkt eitt samstarfsnet (e. coordinating action), DPE, sem snýst um að auka áhuga og vitund fólks um þörf á stafrænni langtímavarðveislu.

Ljóst er að Ísland hefur ekki mikla sérfræðipækkingu fram að færa til slíkra verkefna, þótt vissulega gætu einstakir hugbúnaðarframleiðendur skoðað þátttöku. Það sem Ísland hefur kannski helst fram að færa í slíkum verkefnum er að rafræn stjórnsýsla er almennt með ágætum og að gagnamagn er viðráðanlegt, þannig að mögulegt er innan slíkra verkefna að vinna með „allt“ stafrænt efni eins lands, sem er ekki víða hægt.

4.10 Önnur kerfi

Við lestur efnis og í svörum safna voru nokkur kerfi nefnd sem eru ekki í notkun en söfnin voru að skoða eða höfðu áhuga á að skoða. Þessi kerfi voru skoðuð stuttlega og verður hér gerð grein fyrir þessum kerfum.

Eina íslenska kerfið sem við þekkjum sem gæti nýst við langtímavarðveislu er FDDS kerfið frá Fakta⁷² sem er almennur skjalabrunnur sem notaður hefur verið til að hýsa umtalsverð skjalasöfn, en styður að öðru leyti ekki sérstaklega við langtímavarðveislu. Þetta kerfi má t.d. bera saman við Document Manager frá IBM, sem er kjarni DIAS kerfisins hjá KB í Hollandi.

Fedora (Flexible Extensible Digital Object Repository Architecture)⁷³ er opið efnisvarðveislukerfi (e. content repository system) sem þróað hefur verið í samvinnu Cornell University Information Science og University of Virginia Library. Kerfið styður OAIS staðalinn og vefur efni inn í XML skjöl, en styður að öðru leyti ekki sérstaklega við langtímavarðveislu. Til dæmis notar Fedora Dublin Core lýsigagnastaðalinn sem var þróaður fyrir bókasöfn en þykir ekki nægilega vel útfærður fyrir langtímavarðveislu.

Dspace⁷⁴ er opið efnisvarðveislukerfi, sem er sérstaklega þróað fyrir rannsóknarumhverfi og rannsóknargögn. Kerfið var þróað í samvinnu MIT Libraries og Hewlett-Packard Labs. Dspace styður ekki heldur langtímavarðveislu sérstaklega og skráir einnig lýsigögn í samræmi Dublin Core staðalinn. Kerfið var m.a. notað á safni í Rotterdam, þar sem það fékk þá dóma að það hentaði ekki mjög fyrir stór söfn.⁷⁵

⁶⁸ European Commission. *ICT – Information and communication technologies*. 2006. Bls. 36-39. <ftp://ftp.cordis.lu/pub/fp7/ict/docs/ict-wp-2007-08-draft-ist-2006.pdf>

⁶⁹ <http://www.casparpreserves.eu/>

⁷⁰ <http://www.planets-project.eu/>

⁷¹ <http://www.planets-project.eu/partners/>

⁷² Fakta ehf. <http://www.fakta.is/>

⁷³ <http://www.fedora.info/>

⁷⁴ <http://www.dspace.org/>

⁷⁵ Horsman, P. *Building a Digital Archive*. 2005. http://www.rlg.org/en/page.php?Page_ID=20865#article2

Í Paradigm verkefninu,⁷⁶ sem er samstarfsverkefni bókasafna háskólanna í Oxford og Manchester, voru bæði þessi kerfi metin og talin skorta stóra þætti fyrir langtímarðveislu, svo sem eftirfylgni með skjalasniðum.⁷⁷

Greenstone⁷⁸ er annað opið kerfi fyrir uppbyggingu stafrænna bókasafna, sem þróað var gegnum New Zealand Digital Library Project við University of Waikato í samstarfi við UNESCO. Kerfið skráir lýsigögn í samræmi við Dublin Core staðalinn, en styður langtímarðveislu ekki sérstaklega.

DigiTool⁷⁹ er enn eitt kerfið sem þróað hefur verið fyrir uppbyggingu stafrænna bókasafna. Það er framleitt af Ex Libris, og hefur m.a. verið notað á landsbókasafni Austurríkis. Það styður heldur ekki langtímarðveislu sérstaklega.

Síðasta kerfið sem hér verður nefnt er LOCKSS (Lots Of Copies Keep Stuff Safe).⁸⁰ Þetta er eina opna kerfið sem gerir tilraun til að styðja langtímarðveislu, með því annars vegar að geyma mörg eintök af hverju skjali og nota þau til að komast að því hvenær eintök hafa skemmt, og hins vegar að aðstoða við umbreytingu á gagnasniði (þó í upphafi aðeins fyrir vefefni).⁸¹

⁷⁶ <http://www.paradigm.ac.uk/>

⁷⁷ <http://www.dpconline.org/graphics/join/paradigm.html>

⁷⁸ <http://www.greenstone.org/cgi-bin/library>

⁷⁹ <http://www.exlibrisgroup.com/digitool.htm>

⁸⁰ <http://www.lockss.org/>

⁸¹ Rosenthal, D.S.H. *Transparent Format Migration of Preserved Web Content*. 2005.
<http://www.dlib.org/dlib/january05/rosenthal/01rosenthal.html>

5 Staðlar

Í þessum kafla verður farið stuttlega yfir helstu staðla sem skipta máli við langtímarvarðveislu stafræns efnis. Stöðlum þessum má skipta gróflega í þrjá þætti: Staðla um ferli efnis í slíkum kerfum, staðla um lýsigögn sem nauðsynleg eru fyrir langtímarvarðveislu og staðla um skjalasnið. Hér á eftir verður rætt um hvern þessarra þátta fyrir sig.

5.1 Ferli

Eini staðallinn sem raunverulega skiptir máli varðandi ferli efnis í langtímarvarðveislu er OAIIS staðallinn. Honum var lýst stuttlega að ofan í kafla 3.1 og verður sú lýsing ekki endurtekin. En hér má ítreka að „allir“ nota þennan staðal,⁸² þ.e. öll kerfi sem við höfum rekist á í þessari vinnu. OAIIS staðallinn er nokkuð lauslega skilgreindur og því er hægt að útfæra marga hluti á mismunandi vegu, en staðallinn er fyrst og fremst notaður til að tryggja að kerfin geti gert allt sem gera þarf.

Annar staðall sem nefndur hefur verið í tengslum við langtímarvarðveislu stafræns efnis er MoReq staðallinn.⁸³ Þessi staðall á þó fyrst og fremst við um skjalakerfi fyrir skjöl í vinnslu, þ.e.a.s. kerfi sem vinna með efni áður en því er komið til langtímarvarðveislu. Það má segja að ef notuð eru kerfi sem uppfylla MoReq, þá sé það efni sem úr þeim kemur vel statt hvað varðar langtímarvarðveislu og uppfyllingu OAIIS staðalsins. Þess má geta að Hugvit hyggst styðja MoReq staðalinn í framtíðinni í þeim kerfum þeirra sem notuð eru í stjórnsýslunni í dag.

5.2 Lýsigögn

Eins og áður hefur verið komið inn á, eru lýsigögn nauðsynleg fyrir langtímarvarðveislu stafræns efnis. Lýsigögn lýsa tilurð efnisins og þróun þess og gera notendum framtíðarinnar kleift að treysta efninu, ef rétt er á málum haldið.

Margir lýsigagnastaðlar hafa verið gerðir með það að markmiði að styðja langtímarvarðveislu. Nokkur sátt virðist vera að verða um að lýsigögnin sem PREMIS hópurinn skilgreindi séu þeirra best.⁸⁴ The National Archives, Bretlandi, notar yfirmengi af PREMIS, National Library of Australia notar PREMIS auk smá viðbóta, Koninklijke Bibliotheek notar ekki PREMIS eins og er, en notar skilgreininguna við þróun á næstu útgáfu síns kerfis. Aðrir hafa skoðað PREMIS og borið saman við eigin lýsigögn, s.s. National Archives of Australia sem notar eigin lýsigagnasnið.

Í vinnu PREMIS hópsins var lögð áhersla á að lýsigögnin næðu yfir helstu lykilþarfir fyrir langtímarvarðveislu, innihéldu lýsigögn sem væru mikilvæg fyrir allar tegundir efnis, og væru útfæranleg. Þótt PREMIS lýsigögn séu ekki bundin útfærsluleiðum, er hægt að útfæra þau í XML með fimm skemum. PREMIS er ekki formlegur staðall, en er gerður af vinnuhópi sérfræðinga frá mörgum lykilaðilum í stafrænni varðveislu, og byggir á reynslu af fyrri stöðlum. PREMIS er núna haldið við af Library of Congress.⁸⁵

METS er annar mikilvægur óformlegur „staðall“⁸⁶ sem lýsir hvernig tengja má saman stafrænt efni og lýsigögn þess.⁸⁷ Hann er óháður skjalasniðum og lýsigögnum, og því má

⁸² Thingholm, F. *Kort introduktion til OAIIS*. 2006.

http://www.skjalasafn.is/Syningar/img/Kort_introduktion_til_OAIS_Frank-Thingholm.ppt

⁸³ *MoReq specification*. 2001. <http://www.cornwell.co.uk/moreqdocs/moreq.pdf>

⁸⁴ Verheul, I. *Networking for digital preservation*. 2006. Bls. 48

⁸⁵ <http://www.loc.gov/standards/premis/>

⁸⁶ Lavoie, B. *Preservation metadata*. 2005. <http://www.dpconline.org/docs/reports/dpctw05-01.pdf>

⁸⁷ <http://www.loc.gov/standards/mets/>

nota hann með PREMIS. Staðallinn er þróaður undir handleiðslu Library of Congress í Bandaríkjunum. National Library of Australia notar METS í sínum kerfum, en önnur helstu lönd sem hér eru nefnd hafa ekki nefnt að þau noti METS.

Þess má geta að PDF/A skjöl, og ýmsar aðrar skjalategundir, geta innifalið lýsigögn á XMP sniði (eXtensible Metadata Platform),⁸⁸ sem er byggt á XML. Það snið hefur þann kost að skjölin eru sjálfstæð og læsileg ein sér. Geymsla lýsigagna inni í skjali hefur samt nokkra galla sem vega þarf og meta. Í fyrsta lagi eru ekki allar skjalategundir búnar XMP, sér í lagi er ekki ljóst hvort hægt sé að lýsa gagnagrunnum þannig. Í öðru lagi er ekki hægt að sækja lýsigögnin nema að sækja viðkomandi skjal fyrst. Ýmsir aðilar (t.d. Koninklijke Bibliotheek og The National Archives) telja óskynsamlegt að geyma lýsigögn og skjöl saman, þar sem það geri viðhald lýsigagna miklu erfiðara. Þetta á kannski sérstaklega við á meðan útfærsla lýsigagna er ekki endanlega frágengin og oft er verið að uppfæra lýsigögn eftir því sem kröfurnar breytast.

METS býður aftur á móti upp á báðar aðferðir: Annars vegar að tengja lýsigögn og efni saman í eina heild, og hins vegar að geyma lýsigögnin aðskilin frá viðkomandi efni með vísun á milli. Loks má nefna að XMP sniðið er í eigu Adobe, meðan METS er opinn staðall.

5.3 Skjalasnið

5.3.1 Val á skjalasniðum

Val á skjalasniðum skiptir miklu máli um varðveisluhæfni gagna. Brown⁸⁹ hefur gefið út „Guidance Note“ um val á skjalasniðum. Clausen⁹⁰ lýsir einnig matsaðferð fyrir framtíðarhæfni skjalasniða. Í báðum tilfellum eru opin (e. open), algeng (e. ubiquitous, widespread), stöðug (e. stable), einföld snið talin líklegust til að varðveitast vel. Snið sem eru flókin eða bjóða upp á aðgerðir (e. functionality) eru líklegri til að valda vandræðum. Dulkóðun og þjöppun eru almennt talin til vandræða.

Danska þjóðskjalasafnið hefur gefið út lista, sem þýddur hefur verið í regludrögum ÞÍ. Þar má nefna að skjölum ber að skila sem TIFF myndum, en að auki þarf að skila textanum á textasniði. Í þeim tilfellum þar sem textinn fylgir, er talið ásættanlegt að þjappa myndunum til að spara diskapláss, þótt almennt sé þjöppun talin hefta langtímavarðveislu.

Open Document Format⁹¹ er nýtt staðlað snið (varð ISO/IEC 26300 staðall í maí 2006), byggt á XML, til að geyma efni á almennan hátt. Þetta snið er hins vegar nýtt og því ekki komið í almenna notkun. Sama má segja um XPS (XMP Paper Specification) skjalasniðið,⁹² sem Microsoft hyggst gera að alþjóðlegum staðli.

Umtalsverð vinna hefur farið fram, og mun fara fram, við skráningu skjalasniða og viðhald upplýsinga um þau. T.d. hefur National Archives rekið PRONOM skjalasniðavefinn⁹³ um hríð. Í Harvard hafa menn stungið upp á að búa til net slíkra skjalasniðavefa.⁹⁴

⁸⁸ Wikipedia. *Extensible metadata platform*. Skoðað 08.12.2006.

http://en.wikipedia.org/wiki/Extensible_Metadata_Platform.

⁸⁹ Brown, A. *Preserving the digital heritage*. 2003. <http://www.nationalarchives.gov.uk/documents/brown.pdf>

⁹⁰ Clausen, L.R. *Handling file formats*. 2004. <http://netarchive.dk/publikationer/FileFormats-2004.pdf> Bls. 11-13.

⁹¹ <http://www.odfalliance.org/>

⁹² Microsoft. *XML paper specification: overview*. Skoðað 08.12.2006.

<http://www.microsoft.com/whdc/xps/default.mspx>

⁹³ The National archives. *PRONOM*. Skoðað 08.12.2006. <http://www.nationalarchives.gov.uk/pronom/>

⁹⁴ Abrams, S et al. *A proposal for a global digital format registry*. 2005.

<http://hul.harvard.edu/gdfr/documents/Proposal-2005-09-29.doc>

5.3.2 PDF/A

PDF/A er alþjóðlegur staðall ISO 19005, sem var samþykktur 2005.⁹⁵ PDF/A byggir á PDF 1.4 staðlinum sem Adobe á, en hefur birt og þar með leyft öðrum aðilum að búa til hugbúnað til að lesa og skrifa PDF skjöl. PDF/A staðallinn takmarkar PDF skjöl við þau sem standa alfarið sjálf, þ.e. innihalda stafagerðir og annað sem þarf til að birta skjalið, og bjóða ekki upp á neinar forritaðar aðgerðir.

PDF skjöl geta innihaldið lýsigögn á XMP sniði, sjá kafla um lýsigögn að ofan. Þau lýsigögn eru ekki ákveðin í staðlinum, en eru sveigjanleg og virðast geta ráðið við lýsigagnastaðla.

Skiptar skoðanir eru um það hvort PDF/A teljist opinn staðall, þar sem hann byggir á staðli í eigu Adobe. Í því samhengi má reyndar nefna að TIFF sniðið er einnig í eigu Adobe, en ekki alþjóðlegur staðall. Hins vegar hefur TIFF sniðið verið óbreytt frá 1992 og er útbreitt, þótt hugbúnaði fari fækkandi sem styðji sniðið; t.d. geta vefvafrar almennt ekki birt TIFF.

Bent hefur verið á að fá hugbúnaðarkerfi styðji PDF/A enn sem komið er, vegna þess hversu nýr hann er, og það standi einna helst í vegi fyrir upptöku hans. Þó hafa nokkur skjalasöfn, eins og Riksarkivet í Svíþjóð og NARA í Bandaríkjunum, þegar samþykkt þennan staðal. Við fyrstu sýn virðist skýrsluhöfundum PDF/A vera ásættanlegt varðveislusnið, en nánari skoðun á hvort og hvenær sé raunhæft að taka það upp er utan efnisumfangs þessarar skýrslu.

⁹⁵ [PDF/A]: draft. 2003. [http://www.aiim.org/documents/standards/ISO_19005-1_\(E\).doc](http://www.aiim.org/documents/standards/ISO_19005-1_(E).doc)

6 Samantekt

6.1 Fjármagn

Sérlega mikilvægt er að gæta þess að tryggja nauðsynlegt fjármagn til verkefnis af þessari gerð. Ekki má líta á langtímavarðveislu gagna sem spennandi átaksverkefni, heldur sem nauðsynlegt langtímaverkefni. Vissulega getur þurft átaksfjármagn til að smíða eða kaupa tölvukerfi, þjálfa starfsfólk, og bjarga eldra efni frá glötun. Neyðaraðgerðir eru venjulega merki um fyrri vanrækslu, eru oft mjög kostnaðarsamar, og eru til þess fallnar að veikja trú fólks á virkni varðveislukerfisins. Þær ber því að almennt að forðast en í ljósi þess að langtímavarsla hefur ekki hafist á Íslandi í nokkrum mæli er líklegt að slíkar aðgerðir þurfi.

En langtímavarðveisla krefst mikillar og jafnrar vinnu um ókomna framtíð, og mikilvægt er að ríkisvaldið tryggi nægilegt fjármagn á fjárlögum hvers árs til að sinna því verkefni.

6.2 Lagaleg atriði

Lög og reglur stjórnvalda eru til þess að ná ákveðnum markmiðum sem menn hafa komið sér saman um. Það er því samdóma álit allra viðmælenda og ritaðra heimilda að markvisst lagaumhverfi auðveldar óneitanlega vinnu við langtímavarðveislu skjala. Með skýrum skilaskyldulögum, og sterkum reglugerðum útgefnum af, t.d., forsætisráðuneyti, eru söfnin mun betur í stakk búin til að gera kröfu á skjalamyndara að vinna með söfnunum að varðveislu.

Nú gætu andstæðingar markviss lagaumhverfis bent á að eitt af þeim löndum sem lengst er komið í langtímavarðveislu opins efnis, Holland, hefur ekki skilaskyldureglur fyrir landsbókasafn sitt (KB). Fyrir vikið hefur bókasafnið getað valið sér samstarfsaðila og einbeitt sér að varðveislu tímarita sem gefin eru út af öflugum útgefendum sem sjá sér hag í samvinnu við bókasafnið. Þetta hefur eflaust einfaldað þeirra vinnu, en eftir stendur að annað útgefið efni er ekki varðveitt og safnið vantar skilaskyldureglur til að geta krafist skila.

Árið 2002 stóð þjóðskjalasafn Danmerkur fyrir könnun á hindrunum fyrir langtímavarðveislu skjala, þar sem ráðuneytum og stofnunum var gert að fara í gegnum regluverk sitt og finna slíkar hindranir. Svo var hvert tilvik metið og ákveðið eftir atvikum að 1) fjarlægja hindrunina með breytingum á lögum og/eða reglum, eða 2) að hindrunin væri réttmæt og ekki ætti að fjarlægja hana. Án þess að við höfum getað kynnt okkur nánar hvernig þessi könnun var framkvæmd, þá virðist nauðsynlegt að fara í gegnum slíka vinnu einnig hér.

Almennt má segja að sterk lög um langtímavarðveislu stafrænna gagna, og sér í lagi aðgengi að þeim, leggi þyngri byrðar á söfn; en jafnframt virðast þau lönd sem búa við sterkara lagaumhverfi lengra á veg komin. Eftir að hafa farið yfir helstu lög sem koma að varðveislu stafrænna gagna á Íslandi standa nokkur atriði og spurningar upp úr sem mælt er með að hugað verði sérstaklega að:

- Ábyrgð stofnana á aðgengi að efni í 30 ár:
 - Hvað með tæknilegan búnað og hæft starfsfólk?
 - Hvað með flutning efnis milli kerfa, geta stofnanir séð um slíkan flutning?Hér þarf eflaust að skoða betur hversu lengi stofnun getur staðið undir því að halda utan um stafrænt efni.
- Kröfur vegna afhendingar:
 - eru kröfur Lbs-Hbs og ÞÍ réttar og nægilegar?
 - Nauðsynlegt er að móta grisjunarstefnu til að minnka gagnamagn. Einnig er nauðsynlegt að veita stofnunum ráðgjöf um skjalastjórnun.

- Landsbókasafni-Háskólabókasafni ber að taka við efni á því formi sem þau birtast notendum. Þetta flækir málin verulega þar sem efnið er á mjög breytilegu sniði. Safnið hefur leyfi til að semja um þetta, en ef útgefendur vilja það ekki þá verður safnið að taka við efninu.

Hér þarf að styrkja lagalega stöðu safnanna gagnvart afhendingaraðilum.

- Upplýsingalög og höfundaréttarlög flækja verulega aðgengismál og afhendingar úr söfnunum, ákvæði þeirra verða að vera uppfyllt.
 - Það þarf að standa vel að skráningu efnis í varðveislukerfum með aðgangsmörkingum.
 - Það þarf að tryggja öryggi kerfanna, bæði búnaðarins og efnisins.
 - Auk þess setja upplýsingalögin tímatakmörkun á afgreiðslu á efni úr söfnum, sem kallar á auðvelt aðgengi starfsfólks safnanna að efninu.
 - Það þarf að huga vel að vinnuferlum í söfnunum vegna afhendingar úr söfnunum.
- Sett eru ákveðin tímatakmörk um hvenær opna megi aðgang að efni í söfnum:
 - Efni sem ekki fellur undir takmarkaðan aðgang skv. upplýsingalögum má opna eftir 30 ár, að frátöldu efni um einkamálefni einstaklinga, þá 80 ár.
 - Höfundaréttur gildir í 70 ár. Í Evrópu er verið að reyna að semja almennt við samtök handhafa höfundarréttar til að stytta það tímabil sem efni er lokað vegna höfundarréttar.
- Stjórnsýslulögin setja kröfur um tækni og öryggi, efni skal vera aðgengilegt notendum og sannreynanlegt:
 - Þetta hefur mikil áhrif á langtímavarðveislu stafræns efnis, þessi ákvæði skal uppfylla.

Langtímavarsla stafræns efnis snýst því að hluta um þekkingu og reynslu, auk skýrra og markvissra lagaákvæða. En að stórum hluta er það upplýsingatæknilegt verkefni, þar sem koma þarf upp nauðsynlegum hug- og vélbúnaði til að hægt sé að varðveita efnið til frambúðar.

6.3 Tæknileg atriði

Þegar lítið er til þeirra safna sem lýst er hér að ofan kemur sterk sameiginleg högun í ljós.

- Öll söfnin nota OAIS staðalinn til að stýra ferli stafræns efnis.
- Öll söfnin nota lýsigögn (gjarnan PREMIS) til að lýsa þeim og leggja mikið upp úr gæðum lýsigagnanna.
 - Mörg söfn hafa þó gert eigin viðbætur eða búið til eigin lýsigagnastaðla áður en PREMIS varð til.
- Undirliggjandi varðveislukerfunum er diskastæða og gjarnan bandstöðvar.
 - Skjalasöfnin hafa flest sett upp fleiri en eitt jafngilt kerfi til að auka líkur á vel heppnaðri varðveislu. Bókasöfnin hafa hins vegar látið eitt kerfi og venjulega afritatöku duga, þótt Koninklijke Bibliotheek í Hollandi hyggi á uppsetningu annars kerfis.
 - Sums staðar eru kerfi að keyra á mismunandi vélbúnaði og stýrikerfum, til að minnka líkur á gagnatapi.
 - Aðeins Statens Arkiver í Danmörku notar aðra miðla en segulmiðla, en safnið notar einnig ljósmiðla.

- Mismunandi kerfi eru notuð til að stýra varðveislunni.
 - Sum söfn hafa skrifað eigin hugbúnað, en það er venjulega vegna þess að nægilega góðu hugbúnaður hefur ekki verið til.
 - Almennt telja söfnin að betra sé að kaupa að hugbúnað.
- Flest söfnin geyma mjög viðkvæmt efni í kerfi sem er aðskilið frá umheiminum
 - Aðeins Bretland hefur lagt í nauðsynlegan kostnað til að gera einnig viðkvæmt efni aðgengilegt út á við gegnum öruggt net stjórnsýslunnar þar í landi.
- Almennt eru lýsigögn fyrir leit og opið efni geymd í öðru kerfi, sem er sérhannað fyrir leitina.
- Almennt eru löndin skammt á veg komin í virkri langtímavarðveislu.
 - Statens Arkiver og National Archives of Australia hafa skilgreint varðveislusnið sem efni er fært á.
 - Koninklijke Bibliotheek hefur hafið vinnu við kerfi sem fylgist með birtingarleiðum og varar við ef eitthvað efni er líklegt til að týnast.

6.4 Niðurstaða

Þegar litið er til tæknilegra lausna þeirra safna sem að ofan er lýst, er það álit skýrsluhöfunda að The National Archives í Bretlandi og Koninklijke Bibliotheek í Hollandi séu bestu fordæmin í nágrannalöndum okkar í dag.

The National Archives hefur skilgreint mjög heilsteypta, sérsmíðaða högun fyrir langtímavarðveislu. Einnig er tenging kerfisins við lokað net stjórnsýslunnar mjög áhugaverð.

Koninklijke Bibliotheek hefur einnig komið upp áhugaverðu kerfi. Athyglisverðust er þó vinna þeirra við eftirfylgni með skjalasniðum, til að geta tryggt að ávallt séu tvær birtingarleiðir fyrir hvert skjalasnið.

Þótt Ástralir hafi staðið sig vel eru þeir langt í burtu og tímamismunurinn sérlega óhagstæður. Það verður áhugavert að fylgjast með framgangi Svía á næstu árum, þegar LDB-verkefnið fer að skila árangri, og má segja að það eitt og sér sé gott fordæmi. Loks verður áhugavert fyrir Ísland að reyna að tengjast evrópsku rannsóknarverkefnum Planets og CASPAR eða þeim nýju verkefnum sem fara munu af stað undir 7. rammaáætluninni.

7 Heimildaskrá

1. Abrams, S et al. *A proposal for a global digital format registry*. Harvard University Library, Cambridge, USA. 2005. <http://hul.harvard.edu/gdfr/documents/Proposal-2005-09-29.doc>
2. Atkinson, Steve. *Storage management and backup schemes for broadcast video*. Fyrirlestur á ráðstefnu „SMPTE Technical Conference and Exhibition“ í Kaliforníu, Bandaríkjunum, 20.-23. október 2004. http://www.sgluk.com/downloads/docs/storage_management_lo_lo.pdf
3. Bradley, Kevin. *Risks associated with the use of recordable CDs and DVDs as reliable storage media in archival collections: strategies and alternatives*. Unesco, Paris. 2006. <http://unesdoc.unesco.org/images/0014/001477/147782E.pdf>
4. Brown, Adrian. *Preserving the digital heritage: building a digital archive for UK government records*. 2003. Í: Online Information 2003: Proceedings. <http://www.nationalarchives.gov.uk/documents/brown.pdf>
5. Brown, Adrian. *Selecting storage media for long-term preservation*. 2003. (The national archives: digital preservation guidance note ; 2) http://www.nationalarchives.gov.uk/documents/selecting_storage_media.pdf
6. CASPAR. <http://www.casparpreserves.eu/>
7. Clausen, Lars R. *Handling file formats*. Århus, 2004. <http://netarchive.dk/publikationer/FileFormats-2004.pdf>
8. Consultative committee for space data systems. *Reference model for an open archival information system : blue book*. 2002. <http://public.ccsds.org/publications/archive/650x0b1.pdf>
9. Department for constitutional affairs. *Freedom of information*. Skoðað 29.10.2006. <http://www.dca.gov.uk/foi/index.htm>
10. DigiTool. <http://www.exlibrisgroup.com/digitool.htm>
11. Dspace. <http://www.dspace.org/>
12. Dunn, J.R. *The EMP threat: ElectroMagnetic Pulse warfare*. 2006. Skoðað 13.11.2006. <http://www.leatherneck.com/forums/archive/index.php/t-28864.html>
13. Edelholt, Peter. Fyrirlestur á ráðstefnu „Seminar om tilgængeliggørelse af elektroniske arkivalier“ haldin á Íslandi 25.-26. apríl 2006. http://www.skjalasafn.is/Syningar/img/Peter_Edelholt.ppt
14. eGovernment News. *UK: Improved version of the UK Government Secure Intranet launched*. 2004. <http://ec.europa.eu/idabc/en/document/2136>
15. European Commission. *ICT – Information and communication technologies. Work programme 2007-08. Draft*. 2006. <ftp://ftp.cordis.lu/pub/fp7/ict/docs/ict-wp-2007-08-draft-ist-2006.pdf>
16. Fakta ehf. <http://www.fakta.is/>
17. Fedora. <http://www.fedora.info/>
18. FileTek. *FileTek Successfully Writes One Billion Files to a Single StorHouse/RFS File System : press release*. 2005. Skoðað 29.01.2007. http://www.filetek.com/press/releases/2005/pr_one_billion_files.htm

19. Forsætisráðuneytið. *Skýrsla nefndar um rafræna stjórnýslu*. Reykjavík, 2002.
20. Gladney, H.M. *Digital preservation*. Í: Digital document quarterly, vol. 5 no. 2. 2005 http://home.pacbell.net/hgladney/ddq_5_2.htm
21. Greenstone digital library software. <http://www.greenstone.org/cgi-bin/library>
22. Horsman, P. *Building a Digital Archive: a Dutch experience*. Í: RLG DigiNews. Vol. 9, no. 6. 2005. http://www.rlg.org/en/page.php?Page_ID=20865#article2
23. *Höfundalög nr. 73/1972*
24. *i2010: Digital libraries initiative*. http://ec.europa.eu/information_society/activities/digital_libraries/index_en.htm
25. JaJa, Joseph, et al. *Digital archiving and long term preservation*. The national archives, Bandaríkin, 2004. Skoðað 13.11.2006. <http://www.archives.gov/era/papers/thic-04.html>
26. Jespersen, Lone Smith. *Tilgængeliggjørelse af elektroniske arkivalier: resultatkravprojekt 2005-2006*. Fyrirlestur á ráðstefnu „Seminar om tilgængeliggjørelse af elektroniske arkivalier“ haldin á Íslandi 25.-26. apríl 2006. http://www.skjalasafn.is/Syningar/img/Projektpraesentation_LSJ.ppt
27. Joint information systems committee. <http://www.jisc.ac.uk/>
28. Koninklijke Bibliotheek. *Web archiving*. Skoðað 29.10.2006. http://www.kb.nl/hrd/dd/dd_projecten/projecten_webarchivering-en.html
29. KOPAL. <http://kopal.langzeitarchivierung.de/index.php.en>
30. Lavoie, Brian, et al. *Preservation metadata: technology watch report*. Digital preservation coalition, 2005. <http://www.dpconline.org/docs/reports/dpctw05-01.pdf>
31. LDB. <http://ldb.project.ltu.se/>
32. Library and Archives Canada. *Toward a Canadian digital information strategy*. Skoðað 29.10.2006. <http://www.collectionscanada.ca/scin/index-e.html>
33. LOCKSS. <http://www.lockss.org/>
34. *Lög um Landsbókasafn Íslands – Háskólabókasafn nr. 71/1994*.
35. *Lög um skylduskil til safna nr. 20/2002*.
36. *Lög um Þjóðskjalasafn Íslands nr. 66/1985*.
37. Madsen, Kirsten. *Debatoplæg om benyttelse af de elektroniske arkivalier*. Fyrirlestur á ráðstefnu „Seminar om tilgængeliggjørelse af elektroniske arkivalier“ haldin á Íslandi 25.-26. apríl 2006. http://www.skjalasafn.is/Syningar/img/Debatoplaegisland_KM.ppt
38. Menntamálaráðuneytið. *Skýrsla nefndar um varðveislu tölvugagna sem verða til í stjórnýslunni*. 1998. http://www.skjalasafn.is/Syningar/img/Skyrsla_nefndar_um_rafraen_skjol_1998.pdf
39. METS. <http://www.loc.gov/standards/mets/>
40. Microsoft. *XML paper specification: overview*. Skoðað 08.12.2006. <http://www.microsoft.com/whdc/xps/default.msp>
41. *MoReq specification*. 2001. <http://www.cornwell.co.uk/moreqdocs/moreq.pdf>

42. The National archives. *ERA documentation*. Skoðað 13.11.2006.
<http://www.archives.gov/era/about/documentation.html>
43. The National archives. *National archives names Lockheed Martin to build archives of the future: press release*. 2005. <http://www.archives.gov/press/press-releases/2005/nr05-112.html>
44. The National archives. *PRONOM*. Skoðað 08.12.2006.
<http://www.nationalarchives.gov.uk/pronom/>
45. The National archives. *Seamless flow*. Skoðað 29.10.2006.
http://www.nationalarchives.gov.uk/electronicrecords/seamless_flow/default.htm
46. National archives of Australia. *An approach to preservation of digital records*. 2002.
http://www.naa.gov.au/recordkeeping/er/digital_preservation/Green_Paper.pdf
47. National archives of Australia. *The Commonwealth record series (CRS) system*. Skoðað 29.10.2006. <http://www.naa.gov.au/fsheets/fs06.html>
48. National archives of Australia. *Digital preservation: digital preservation software applications*. Skoðað 29.10.2006.
<http://www.naa.gov.au/recordkeeping/preservation/digital/applications.html>
49. National archives of Australia. *Digital preservation: illuminating the past, guiding the future*. 2006.
http://www.naa.gov.au/recordkeeping/preservation/digital/XENA_brochure.pdf
50. National archives of Australia. *Digital preservation: XML data formats*. Skoðað 29.10.2006.
http://www.naa.gov.au/recordkeeping/preservation/digital/xml_data_formats.html
51. National archives of Australia. *Digital preservation project*. [An animation]. Skoðað 29.10.2006.
<http://www.naa.gov.au/recordkeeping/preservation/digital/animation/index.html>.
52. National archives of Australia. *Records and the public*. Skoðað 29.10.2006.
http://www.naa.gov.au/recordkeeping/access/records_public.html
53. National library of Australia. *Preserving access to digital information*. Skoðað 29.10.2006. <http://www.nla.gov.au/padi/>
54. NDIIPP. <http://www.digitalpreservation.gov/>
55. Nielsen, Anders Bo, et al. *Preservation of Electronic Records: experiences from Denmark*. Fyrirlestur á ráðstefnu „International Congress on Archives“ haldin í Vín, Austurríki, 23.-29. ágúst 2004.
http://www.wien2004.ica.org/imagesUpload/pres_190_NIELSEN_A_DEN01E.pdf
56. ODF Alliance. <http://www.odfalliance.org/>
57. Olsen, René Mittå. *Filformater*. Statens arkiver, Danmark. 2000.
http://kum.inforce.dk/graphics/kum/billeder/temaer/tema_bevaring/bilag_til_bevaring_af_kulturarven/bilag_13.doc
58. Oltmans, Erik, van Diessen, Raymond og van Wijngaarden, Hilde. *Preservation functionality in a digital archive*. 2004. Í: Proceedings of the 4th ACM/IEEE-CS joint conference on digital libraries.
59. Palm, Jonas. *The digital black hole*. Riksarkivet, Stockholm, 2006? http://www.tape-online.net/docs/palm_black_hole.pdf

60. PANDORA. <http://pandora.nla.gov.au/index.html>
61. Paradigm. <http://www.paradigm.ac.uk/>
62. [PDF/A]: draft. 2003. [http://www.aiim.org/documents/standards/ISO_19005-1_\(E\).doc](http://www.aiim.org/documents/standards/ISO_19005-1_(E).doc)
63. Picture Australia. <http://www.pictureaustralia.org/>
64. PLANETS. <http://www.planets-project.eu/>
65. PREMIS. <http://www.loc.gov/standards/premis/>
66. PREMIS. *Data dictionary for preservation metadata*. OCLC, 2005. <http://www.oclc.org/research/projects/pmwg/premis-final.pdf>
67. RLG. *Trusted digital repositories: draft for public comment*. Kalifornía, 2005. <http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf>
68. Rosenthal, David S.H., et al. *Transparent Format Migration of Preserved Web Content*. Í: D-Lib Magazine. Vol. 11, no. 1. 2005. <http://www.dlib.org/dlib/january05/rosenthal/01rosenthal.html>
69. Spencer, Jack. *The electromagnetic pulse commission warns of an old threat with a new face*. 2004. Skoðað 13.11.2006. <http://www.heritage.org/research/nationalsecurity/bg1784.cfm>
70. Statens arkiver. *Hvad har kornblomster med CD-R og elektronisk opbevaring af arkivalier at gøre?* Danmark, 1999. <http://www.sa.dk/sa/itogarkiv/teknologi/CDRartikel/ARKIVCD.htm>
71. Steenbakkens, Johan F. *Digital archiving in the twenty-first century: practice at the National Library of the Netherlands*. Í: Library Trends. Vol.54, no. 1. Summer 2005.
72. *Stjórnisýslulög nr. 37/1993*.
73. Tessella support services. <http://www.tessella.com/>
74. Thingholm, Frank. *Kort introduktion til OAIS*. Fyrirlestur á „Seminar om tilgængeliggørelse af elektroniske arkivalier“ haldin á Íslandi 25.-26. apríl 2006. http://www.skjalasafn.is/Syningar/img/Kort_introduktion_til_OAIS_Frank-Thingholm.ppt
75. *Upplýsingalög nr. 50/1996*.
76. Verheul, Ingeborg. *Networking for digital preservation: current practice in 15 national libraries*. K.G. Saur: München, 2006. (IFLA publications : 119).
77. Waller, Martin og Sharpe, Robert. *Mind the gap: assessing the digital preservation needs in the UK*. Digital preservation coalition, 2006. <http://www.dpconline.org/docs/reports/uknamindthegap.pdf>
78. Wikipedia. *Base64*. Skoðað 06.12.2006. <http://en.wikipedia.org/wiki/Base64>
79. Wikipedia. *Extensible metadata platform*. Skoðað 08.12.2006. http://en.wikipedia.org/wiki/Extensible_Metadata_Platform
80. Wikipedia. *Metadata*. Skoðað 13.11.2006. <http://en.wikipedia.org/wiki/Metadata>
81. Þjóðskjalasafn Íslands. *Rafræn skjala- og gagnavarsla ríkisstofnana: könnun Þjóðskjalasafns á skjalavörslu ríkisstofnana árið 2004*. 2005.

http://www.skjalasafn.is/Syningar/img/Rafraen_skjalavarsla_rikisstofnana_2004_lokagerd.pdf

82. Þjóðskjalasafn Íslands. *Reglur Þjóðskjalasafns Íslands um rafræn gagna- og skjalasöfn opinberra aðila. Lokadrög.* 2005. http://www.skjalasafn.is/Syningar/img/Drog_ad_reglum_um_rafraen_skjol_opinber_2005.pdf