

KANSALLISEN DIGITAALISEN KIRJASTON STANDARDISALKKU

24.2.2014



Tämä määrittely on osa opetus- ja kulttuuriministeriön
Kansallinen digitaalinen kirjasto -hanketta (hankenumero OKM052:00/2011)



Lisenssi

Creative Commons Suomi CC-BY-NC-SA 3.0

KDK:n standardisalkun käyttäjälle annetaan oikeus jakaa eli kopioida, levittää, näyttää ja esittää teosta sekä valmistaa muutettuja teoksia seuraavilla ehdolla:

- *Opetus- ja kulttuuriministeriö nimetään teoksen tekijäksi (ei kuitenkaan siten, että ilmoitus viittaisi lisenssinantajan tukevan lisenssinsaajaa tai teoksen käyttötapaa).*
- *Epäkaupallinen osapuoli eli käyttäjä ei saa käyttää teosta kaupallisesti.*
- *Mikäli käyttäjä tekee muutoksia tai käyttää teosta omien teostensa pohjana, tulee johdannaisteos jakaa samalla tai samankaltaisella lisenssillä.*

SISÄLTÖ

1	JOHDANTO.....	5
2	YLEISET STANDARDIT	6
3	METATIEDOT.....	6
3.1	Kuvailusäännöt.....	6
3.2	Kuvaileva metatieto.....	7
3.3	Henkilöiden ja organisaatioiden tiedot.....	7
3.4	Sisällönkuvailuun käytettävät välineet	8
3.5	Hallinnollinen metatieto.....	8
3.6	Rakenteellinen metatieto	8
4	RAJAPINNAT	9
4.1	Metatiedon haravointi.....	9
4.2	Haku.....	9
4.3	Sovellustason rajapinnat.....	10
5	LINKITYSPALVELUT JA TUNNISTEET	10
5.1	Linkitys.....	10
5.2	Tunnisteet: aineisto.....	10
5.3	Tunnisteet: tekijät	10
6	AUTENTIKOINTI JA AUKTORISOINTI	11
6.1	Tiivistealgoritmit.....	11
7	PITKÄAIKAISSÄILYTYS JA AINEISTOJEN SIIRTO	12
8	LÄHTEET	12
9	TÄRKEIMMÄT ASIASANASTOT	13
9.1	Kotimaisia	13
9.1.1	Yleisluontoiset.....	13
9.1.2	Tieteenalakohtaiset, laajahko ala.....	13
9.1.3	Tieteenalakohtaiset, suppeahko ala.....	13
9.2	Ulkomaisia.....	14
9.3	Eurooppalaisia.....	14
10	STANDARDIEN KUVAUKSET	14
10.1	Metatietostandardit.....	14
10.2	Rajapintastandardit.....	15
10.2.1	Metatietojen haravointi	15
10.2.2	Haku	15
10.2.3	Sovellustason rajapinnat	16
10.3	Tunnisteet.....	16
10.4	Pitkäaikaissäilytys ja aineistojen siirto.....	17



Standardisalkku on osa Kansallisen digitaalisen kirjaston kokonaisarkkitehtuuria. Kokonaisuudessaan arkkitehtuuri on julkaistu verkossa: <http://www.kdk.fi/fi/kokonaisarkkitehtuuri>. Standardisalkusta voi antaa palautetta KDK-verkkosivujen palautelomakkeella.

Aluekohtaisen salkun ohella KDK:ssa sovelletaan myös Julkishallinnon standardisalkkua (katso <https://www.yhteentoimivuus.fi/aihealue/Standardit> sekä <http://www.jhs-suositukset.fi/suomi/jhs181>).

1 JOHDANTO

Kansallisen digitaalisen kirjaston standardit määritellään tässä standardisalkussa, joka sisältää tarvittaessa myös lyhyet perustelut valinnoille. Salkku ei anna yksityiskohtaisia ohjeita standardien implementoinnista, vaan ne laaditaan tarvittaessa erikseen. Toisinaan riittää viittaus jo olemassa olevaan kansainväliseen sovellusohjeeseen (profiiliin), mutta myös kansallista ohjeistusta tarvitaan. Aluekohtaisen salkun ohella Kansallisessa digitaalisessa kirjastossa sovelletaan myös Julkishallinnon standardisalkkua¹.

Standardoinnin tavoitteena on varmistaa KDK-järjestelmäkokonaisuuden toimivuus, joka edellyttää esimerkiksi eri organisaatioiden tallentaman metatiedon yhteismitallisuutta. Lisäksi standardoinnilla pyritään takaamaan se, että palvelut kyetään aikanaan siirtämään uusiin laitteisto- ja ohjelmistoympäristöihin mahdollisimman vaivattomasti. Tämä on tärkeää ennen muuta PAS-ratkaisulle.

Käytettävät standardit määritellään standardisalkussa, joka sisältää myös lyhyet perustelut valinnoille. Standardit ovat joko välttämättömiä (must; numeroarvo 1), suositeltavia (should; numeroarvo 2) tai hyödyllisiä (may; numeroarvo 3). KDK-organisaatioiden taustajärjestelmiä on aktiivisesti kehitetty, mutta on mahdollista että jotkin niistä eivät edelleenkaan tue edes kaikkia 1-tason standardeja; tästä aiheutuvat ongelmat tietojen siirrossa ja säilytyksessä ratkaistaan erikseen. Hankittavien uusien järjestelmien on tuettava ainakin välttämättömiä standardeja.

Suoria kansainvälisiä malleja standardisalkun laadintaan ei ole. JISC (Joint Information Systems Committee) on laatinut standardisalkun² joka kattaa useita KDK:n kannalta relevantteja alueita, mutta siitä puuttuu esimerkiksi PAS-ratkaisun ja laajemmin digitaalisen kirjaston edellyttämiä standardeja. Lisäksi JISCillä on informaatioympäristön määrittely³ joka on KDK:n tarpeisiin sovellettavissa. Laadinnassa on käytetty hyväksi myös Minerva-hankkeen julkaisemaa opasta "*Technical guidelines for digital cultural content creation programmes (version 2.0, September 2008)*".

KDK:n standardisalkku sisältää sekä virallisia että ns. teollisuusstandardeja. Lisäksi mukana on standardeja joiden viimeistely on vielä kesken, sillä digitaalisen kirjaston standardipohja on yhä osittain rakenteilla. Kaikki listatut standardit ovat avoimia eli sekä standardit itse että niitä koskeva dokumentaatio on julkisesti käytettävissä⁴. Osa salkun standardeista on maksuttomia, osa maksullisia.

Standardisalkkua täydennetään seuraavasti:

1. Käyttöoikeuksien metatiedosta sekä eri aineistotyyppien (teksti, audio, still-kuva, elävä kuva) teknisestä metatiedosta laaditaan erilliset suositukset.
2. Yksittäisille standardeille laaditaan tarvittaessa soveltamisohjeita eli profiileja ohjelmistokehittäjien ja järjestelmien ylläpitäjien tueksi. Esimerkki tällaisesta profiilista on ohje METS-standardin käytöstä pitkäaikaissäilytysjärjestelmään siirrettävien tietopakettien rakennemäärittelynä^{5,6}.

¹ <https://www.yhteentoimivuus.fi/aihealue/Standardit>, <http://www.jhs-suositukset.fi/suomi/jhs181>

² http://standards-catalogue.ukoln.ac.uk/index/JISC_Standards_Catalogue

³ <http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/standards/>

⁴ Standardien avoimuudesta katso <http://www.jisc.ac.uk/whatwedo/services/techwatch/standards/open.aspx>

⁵ <http://www.kdk.fi/fi/pitkaaikaissailytys/maeaerittely-ja-dokumentit/5-suomi/pitkaaikaissaelytys/145-kdkn-hallinnolliset-ja-rakenteelliset-metatiedot-ja-aineiston-paketointi>

⁶ METS-profiileista yleensä katso <http://www.loc.gov/standards/mets/mets-profiles.html>

Standardisalkun implementointi voidaan vaiheistaa: asiakasliittymää koskevat vaatimukset viimeistään palvelun käyttöönottoon mennessä, PAS-ratkaisun vaatimukset vastaavasti kokonaisuudessaan palvelun käyttöönottoon mennessä, mutta siirtopaketin osalta mahdollisimman nopeasti.

KDK-hanke suunnittelee ja organisoii salkkuun liittyvän koulutuksen. Salkun ylläpidosta vastaa Kansallinen digitaalinen kirjasto -hankkeen ohjausryhmä.

2 YLEISET STANDARDIT

Asiakasliittymän ja PAS-järjestelmän on tuettava UNICODE-standardin⁷ versiota 6.X tai uudempaa (UTF-8 -koodauksella) (1).

Viitetieto ja tekstidokumentit on suositeltavaa toimittaa taustajärjestelmistä tässä merkistössä (2).

Asiakasliittymän ja PAS-järjestelmän käyttöliittymien on tuettava Web Content Accessibility Guidelines (WCAG) -suosituksen versiota 2.0⁸ tason A vaatimusten mukaisesti⁹ (1).

Viranomaisten asiakirja-aineistojen on välttämätöntä noudattaa arkistolaitoksen Sähke-määräyksessä määritellyjä tiedostomuotoja (1).

3 METATIEDOT

Sekä asiakasliittymän että PAS-ratkaisun toiminta edellyttää riittäviä ja sisällöllisesti yhteneviä metatietoja. Siksi metatietoformaateista ja kuvailun periaatteista sopiminen (kuvailvan metatiedon osalta) on KDK:n kannalta välttämätöntä.

Metatieto jakautuu perinteiseen kuvailevaan sekä hallinnolliseen metatietoon. Tällä hetkellä erityyppisille metatiedoille sovelletaan eri formaatteja, joiden tietoelementit ovat osittain päällekkäisiä¹⁰.

3.1 Kuvailusäännöt

Kuvailun kehittämisen perustaksi suositellaan joko RDA- (Resource Description and Access)¹¹ tai Spectrum-säännöstöä¹², joista jälkimmäistä sovelletaan museosektorilla (2). Näiden säännöstöjen soveltamista koordinoidaan siten, että kuvailut ovat Finnan ja PAS-ratkaisun kannalta riittävän yhteismitallisia.

RDA:n lopullinen versio on julkaistu 2011, ja ensimmäiset kirjastot ottavat sen käyttöön vuoden 2013 aikana. Suomalaisen version laadinnasta vastaa Kansalliskirjasto. Kansallisen sovelluksen teossa pyritään ottamaan KDK:n ja eri sektoreiden tarpeet huomioon. Spectrumia käännetään Museo 2015 -hankkeessa, tätä kirjoitettaessa suomennosta viimeistellään.

⁷ <http://unicode.org/>

⁸ <http://www.w3.org/TR/WCAG20>

⁹ <http://www.w3.org/WAI/WCAG20/quickref/>

¹⁰ <http://www.loc.gov/marc/bibliographic/mapping/index.html>

¹¹ <http://www.rdaonline.org/> <http://www.rda-jsc.org/rda.html>

¹² <http://www.collectionslink.org.uk/programmes/spectrum>

3.2 Kuvaileva metatieto

Kukin sektori käyttää kuvailussa yhtä tai useampaa metatietoformaattia. Mahdolliset sovelluskohtaiset formaatit tulisi muuntaa ennen asiakasliittymään siirtämistä johonkin tiedonsiirrossa yleisesti tunnettuun vaihtoformaattiin. Asiakasliittymässä tiedot tallennetaan siinä muodossa, jossa ne on vastaanotettu. Kunkin tietoelementin osalta päätetään erikseen, miten se indeksoidaan. Tämä onnistuu helpoimmin jos kuvailuperiaatteet ja formaattien soveltamistavat ovat riittävän yhtenäiset.

PAS-järjestelmässä kaikki kuvaileva metatieto tallennetaan sellaisenaan pitkäaikaissäilytyksen tueksi.

Kuvailevan metatiedon siirrossa on suositeltavaa soveltaa seuraavia formaatteja (2):

- MARC 21 (MACHINE-Readable Cataloguing)
- FINMARC (siirtymäkauden ajan)
- Dublin Core
- MODS (Metadata Object Description Schema)
- EAD (Encoded Archival Description)
- EAC (Encoded Archival Context)
- LIDO (Lightweight Information Describing Objects)
- VRA Core (Visual Resources Association Core Categories)
- Film identification – Minimum set of metadata for cinematographic works (EN 15744)
- DDI (Data Documentation Initiative)

Perustelut

Jokaisella sektorilla on olemassa yksi tai useampia yleisesti tunnettuja kuvailuformaatteja, joita tulisi käyttää vaihtoformaattina. Jos tämä ei ole mahdollista, metatieto siirretään ja tallennetaan sellaisenaan Finnaan, ja indeksoidaan luovuttajan antaman dokumentaation avulla. Käytännön toteutuksesta sovitaan tarpeen mukaan erikseen.

Dublin Coren osalta sallitaan erilaisten sovellusprofiilien kuten Darwin Coren käyttö, mutta niitä koskevat indeksointiperiaatteet on sovittava asiakasliittymän ylläpitäjäorganisaation kanssa erikseen.

Lisätietoja

Formaatteja vertailemalla on luotu taulukoita siitä, miten formaattien tietoelementit vertautuvat toisiinsa¹³. Normalisointi on haasteellista, jos yhteisiä tietoelementtejä on niukasti tai jos tietojen muoto vaihtelee. Yhteiset kuvailun periaatteet helpottavat tilannetta, mutta sen lisäksi tarvitaan kuvailun yhtenäistämistä eritoten sisällönkuvailun osalta.

3.3 Henkilöiden ja organisaatioiden tiedot

Henkilöiden ja organisaatioiden metatiedot suositellaan tallennettavaksi niin kattavina, että ne mahdollistavat yksiselitteisen tunnistamisen. Yleissanasto- eli YSR-ryhmä on määritellyt joulukuussa 2012 11 tietoelementtiä, jotka on poimittu EU:n Core Person -määrityksestä¹⁴. Tietoihin voi sisältyä esimerkiksi henkilön nimen erilaiset kirjoitusasut, mahdolliset salanimet ja syntymä- ja kuolinvuosi. Organisaatioista on tarpeen tallentaa sekä organisaation nykyinen että edelliset nimet molemmilla kotimaisilla kielillä; valitettavasti edeltävät nimet eivät sisälly EU:n Core Business -määritykseen. Siirrettävän nimitiedon rakenteen määrittelyssä tulisi soveltaa JHS-suositusta 170¹⁵ (2).

¹³ <http://www.oclc.org/research/researchworks/schematrans/default.htm>

¹⁴ http://joinup.ec.europa.eu/asset/core_person/description

¹⁵ <http://www.jhs-suositukset.fi/suomi/jhs170>

3.4 Sisällönkuvailuun käytettävät välineet

Kuvailussa on suositeltavaa käyttää yleisluontoisten tai tieteenalakohtaisten kotimaisten tesaurusten (esim. YSA, FinMeSH) tai vastaavien ontologioiden (kuten YSO:n) termejä (2). Niiden rinnalla/täydennykseksi voidaan aihealueen niin vaatiessa soveltaa kansainvälisiä järjestelmiä.

Sanastojen kuvaamiseen suositellaan SKOS-määrittystä (Simple Knowledge Organization System) (2). Ontologioiden kuvaamisessa sovelletaan edellisen lisäksi RDFS (RDF Schema) ja OWL2 (Web Ontology Language, versio 2) -ontologiakielimäärittäjiä (2).

3.5 Hallinnollinen metatieto

Hallinnollisen metatiedon osalta pyritään kaikilla sektoreilla yhtenäisiin ratkaisuihin. Kyseessä on uusi toiminta-alue, jolla ei ole vakiintuneita käytänteitä. Teknisen metatiedon formaatteja on käytettävissä vain yksi kutakin aineistotyyppiä kohden. Kaikki alla kuvatut määritykset ovat Kongressin kirjaston ylläpitämiä. AudioMD- ja VideoMD-formaatteja on laajennettu kansainvälisenä yhteistyönä 2010-2011.

Tekninen metatieto

Suosittelaa seuraavien formaattien käyttöä (2):

- Kuvat: MIX (NISO Metadata for images in XML Schema)¹⁶
- Teksti: textMD (Technical metadata for Text)¹⁷
- Audio: AudioMD (Technical metadata for Audio)¹⁸
- Video: VideoMD (Technical metadata for Video)

Pitkäaikaissäilytyksen metatieto

Seuraavien formaattien käyttö on välttämätöntä (1):

- PREMIS (Data dictionary for preservation metadata), versio 2.1 tai uudempi.

Käyttöoikeustiedot / Käyttörajoitustiedot

PAS-ratkaisun käyttöoikeustiedot tallennetaan PREMIS-metatietoformaatin¹⁹ Rights-osioon myöhemmin sovittavalla tavalla (1).

Taustajärjestelmiin suositellaan tallennettavaksi aineiston tekijänoikeudellinen tai sopimuksellinen status siten, että tiedot ovat siirrettävissä Finnaan jossakin KDK:n tukemassa vaihtoformaatissa (2).

Lisenssien ja muiden sopimuksien sisältö voidaan tarvittaessa kuvata esimerkiksi ccREL-²⁰ tai ONIX PL –formaatin²¹ avulla (3).

3.6 Rakenteellinen metatieto

Dokumenttien rakenteen kuvaamiseen on käytettävä METS StructMap -standardia (tai vastaavia MXF-standardin ominaisuuksia) (1).

¹⁶ <http://www.loc.gov/standards/mix/>

¹⁷ <http://www.loc.gov/standards/textMD/>

¹⁸ <http://www.loc.gov/standards/amdvmd/>

¹⁹ <http://www.loc.gov/standards/premis/>

²⁰ <http://www.w3.org/Submission/ccREL/>

²¹ <http://www.editeur.org/21/ONIX-PL/>

4 RAJAPINNAT

4.1 Metatiedon haravointi

Kaikkien KDK-hankkeessa mukana olevien tahojen pitää mahdollistaa metatiedon automaattinen haravointi taustajärjestelmistään. Jos tämä ei ole alkuvaiheessa mahdollista, metatiedot voidaan väliaikaisesti koota eräpoimintana ja toimittaa asiakasliittymän ylläpito-organisaatiolle erikseen sovittavalla tavalla (esimerkiksi FTP-siirtona n kertaa vuodessa).

Metatietojen haravointiin sovelletaan OAI-PMH -protokollaa (The Open Archives Initiative Protocol for Metadata Harvesting) versiota 2.0 (1) tai vaihtoehtoisesti Atom-standardia^{22,23} (3).

OAI-PMH:n toteutuksessa on suositeltavaa noudattaa ohjetta Best practices for OAI-PMH DataProvider implementations and shareable metadata²⁴ (2).

OAI-PMH:lla haravoitavien järjestelmien on suositeltavaa tarjota OAI-PMH DC XML -skeeman²⁵ mukaisia tietueita (2). Muiden skeemojen käytöstä sekä mahdollisista laajennuksista tai muutoksista DC XML -skeemaan sovitaan erikseen.

Atom-protokollan käytöstä sovitaan erikseen.

Metatietojen indeksointijärjestelyistä asiakasliittymässä ja mahdollisista normalisointijärjestelyistä Formulaformaattimuuntimessa sovitaan erikseen.

4.2 Haku

Asiakasliittymässä on oltava SRU-palvelin, joka tukee

- SRU-standardin versiota 1.2²⁶ (1) ja
- SRU:n palvelua Search/Retrieve Operation (1).

Asiakasliittymässä on suositeltavaa olla OpenSearch-hakurajapinta²⁷ (2).

Asiakasliittymässä on oltava portaalitoiminnot mahdollistavat Z39.50- ja SRU-asiakasohjelmat (1).

Asiakastietojen hakua ja muokkausta varten taustajärjestelmissä ja asiakasliittymässä on suositeltavaa olla joko 3M Standard Interchange Protocol (SIP2) -rajapintastandardin tai NISO:n Circulation Interchange Protocol (NCIP) -standardin²⁸ tuki (2). Elleivät nämä standardit ole käytettävissä, sovellusten pitäisi tukea Simple Object Access Protocol (SOAP) -standardin versiota 1.2²⁹ tai Representational State Transfer (REST) -määrittystä³⁰ (2).

²² <http://tools.ietf.org/html/rfc4287>

²³ <http://bitworking.org/projects/atom/rfc5023.html>

²⁴ <http://www.diglib.org/pubs/dlf108.pdf>

²⁵ http://www.openarchives.org/OAI/2.0/oai_dc.xsd

²⁶ <http://www.loc.gov/standards/sru/specs/index.html>

²⁷ <http://www.opensearch.org/Home>

²⁸ <http://www.niso.org/workrooms/ncip>

²⁹ <http://www.w3.org/TR/soap/>

³⁰ http://en.wikipedia.org/wiki/Representational_State_Transfer

4.3 Sovellustason rajapinnat

Integroidun kirjastojärjestelmän pitää (1) tarjota käyttöliittymäsovellukselle ja muille KDK:n kannalta relevanteille järjestelmille Digital Library Foundationin ILS-DI -standardin mukainen tiedonsiirto- ja palvelurajapinta³¹. Muut taustajärjestelmät voivat (3) tukea ILS-DI:tä.

5 LINKITYSPALVELUT JA TUNNISTEET

5.1 Linkitys

Linkityspalvelulla tarkoitetaan tunniste- ja muiden metatietojen perusteella tapahtuvaa jatkolinkkien tarjoamista relevantteihin ja käyttäjälle soveltuviin tietoresursseihin. Linkityspalveluissa on käytettävä OpenURL³²-protokollan versiota 0.1 tai versiota 1.0 (1).

Muodostettaessa web-linkkejä digitaalisiin objekteihin ja yksittäisiin tiedostoihin on suositeltavaa käyttää URN-tunnuksiin perustuvia pysyviä web-osoitteita.

5.2 Tunnisteet: aineisto

Julkaistun aineiston identifiointiin tulee käyttää standarditunnuksia, kuten ISBN:ää (International Standard Book Number) tai ISAN:ia (International Standard Audiovisual Number) (1).

Julkaisemattoman aineiston tunnisteena on suositeltavaa käyttää URN:NBN-tunnusta (National Bibliography Number) (2). Organisaatiokohtaisten tunnusten käyttöä on vältettävä, koska ne voivat aiheuttaa ongelmia sekä asiakasliitymässä että PAS-ratkaisussa (kaksi organisaatiota voi antaa saman tunnuksen).

NBN-tunnistejärjestelmän käyttöönotosta on sovittava erikseen Kansalliskirjaston kanssa.

Aineiston yksilöinnissä voidaan käyttää myös URN:OID-tunnuksia. (3). OID on ISO-standardilla määritelty organisaatioiden yksilöintijärjestelmä, jota käyttämällä paitsi organisaatioille ja niiden yksiköille, myös esimerkiksi henkilöille, asiakirjoille ja laitteille voidaan antaa globaalisti yksilölliset tunnisteet. OID-tunnisteiden käytöstä on laadittu julkishallintoa koskeva kansallinen suositus JHS 159³³.

5.3 Tunnisteet: tekijät

Henkilö- ja yhteisötekijöiden tunnistamiseen on suositeltavaa käyttää International Standard Name Identifier (ISNI) –standarditunnuksia (ISO 27729)³⁴ sen jälkeen kun tunnuksen käyttö Suomessa on aloitettu (2). ISNI:n rinnalla voidaan käyttää ORCID-tunnuksia³⁵ (3), joka vastaa rakenteeltaan ISNIä, mutta jolla on eri ylläpito-organisaatio ja jossa tunnisteiden jakeluprosessi on hajautettu, kun se ISNI-järjestelmässä on täysin keskitetty.

Auktoriteettitietokannassa sovellettavasta tunnistejärjestelmästä päätetään erikseen.

³¹ http://old.diglib.org/architectures/ilsdi/DLF_ILS_Discovery_1.0.pdf

³² <http://en.wikipedia.org/wiki/OpenURL>

³³ <http://www.jhs-suositukset.fi/suomi/jhs159>

³⁴ <http://isni.org/>

³⁵ <http://about.orcid.org/>

6 AUTENTIKOINTI JA AUKTORISOINTI

Varmenteiden hankinnassa ja käsittelyssä sekä sähköisiin allekirjoituksiin liittyvien avainten hallinnassa on otettava huomioon laki vahvasta sähköisestä tunnistamisesta ja sähköisistä allekirjoituksista 7.8.2009/617³⁶ ja sen asettamat laatuvaatimukset.

Internetin yli tapahtuvassa käyttäjätunnistuksessa tiedonsiirto salataan Secure Sockets Layer (SSL) v. 3.0 tai SSL:n korvaavalla Transport Layer Security (TLS) -protokollalla (1). Tunnistus voidaan tehdä salaamattomalla yhteydellä vain organisaation sisäisessä lähiverkossa.

Järjestelmien väliseen keskinäiseen tunnistautumiseen tulee käyttää X.509-standardin mukaisia palvelinvarmenteita (1).

Aineiston sähköisessä allekirjoittamisessa tulee käyttää X.509-standardin mukaisia henkilövarmenteita (1).

Heikon Identiteetin tunnistautumispalveluissa tulisi käyttää OpenID:tä (2).

Vahvan identiteetin tunnistautumispalveluilla on oltava käytettävissä julkisen sektorin tunnistautumisfederaatioiden palvelut, esimerkiksi Haka (Shibboleth/SAML) (1), VIRTU (SAML 2.0) (1), VETUMA (SAML 2.0) (1) arjen tietoyhteiskunta -hankkeen tuottaman luottamusverkoston palvelu.

Käyttäjän tunnistamisessa on voitava käyttää LDAP-protokollaa (1).

Aineistojen käyttöoikeuksia voidaan tarvittaessa rajata asiakasliitymässä asiakkaan IP-osoitteen perusteella (2).

Aineistojen käyttöoikeuksia voidaan rajata Shibboleth/SAML auktorisoinnilla perustuen esimerkiksi käyttäjän kotiorganisaatioon ja rooliin (2).

6.1 Tiivistealgoritmit

Tiedostojen eheys on varmistettava tuottamalla niistä mahdollisimman varhaisessa vaiheessa mutta viimeistään ennen PAS-järjestelmään lähettämistä tiiviste. Laskennassa tulee käyttää MD5³⁷-, SHA-1³⁸-, SHA-224³⁹, SHA-256⁴⁰, SHA-384⁴¹ tai SHA-512⁴² algoritmia (1).

³⁶ <http://www.finlex.fi/fi/laki/ajantasa/2009/20090617>

³⁷ <http://fi.wikipedia.org/wiki/MD5>

³⁸ <http://fi.wikipedia.org/wiki/SHA-1>

³⁹ <http://en.wikipedia.org/wiki/SHA-224>

⁴⁰ <http://en.wikipedia.org/wiki/SHA-256>

⁴¹ <http://en.wikipedia.org/wiki/SHA-384>

⁴² <http://en.wikipedia.org/wiki/SHA-512>

7 PITKÄAIKAISSÄILYTYS JA AINEISTOJEN SIIRTO

Pitkäaikaissäilytyksessä sovelletaan OAIS-viitemallia: standardi SFS 5972 (Viitemalli pitkäaikaissäilytysarkistolle) tai vastaava ISO-standardi ISO 14721:2003 (Open Archival Information System) (1).

OAIS-viitemallissa pitkäaikaissäilytykseen siirrettävää yksittäistä aineistokokonaisuutta kutsutaan siirtopaketiksi. Näissä siirtopaketeissa on yhdistettävä sekä itse aineisto että sitä kuvaava hallinnollinen ja kuvaileva metatieto. "Paketointiin" tulee soveltaa METS (Metadata Encoding and Transmission Standard) -standardin⁴³ versiota 1.10 (1), ja METS-pakettien rakenteen tulee olla KDK:n määrittelyn mukainen⁴⁴ (1).

PAS-ratkaisuun siirrettävän aineiston tulee olla Kansallisen digitaalisen kirjaston sallimissa siirto- tai säilytyskelpoisessa tiedostomuodossa⁴⁵ (1). Audiotallenteille voidaan lisäksi soveltaa IASA:n 2009 julkaisemia periaatteita [IASA] ao. tiedostomuotojen osalta (2). Kansallinen digitaalinen kirjasto -hanke valmistelee lisäksi ohjeistusta tiedostoihin liitettävästä teknisestä ja kuvailevasta metatiedosta.

Tiedostojen on oltava siirtopaketissa siirto- tai säilytyskelpoisessa muodossa, mutta ne voidaan sisällyttää pakettiin myös alkuperäisessä tiedostomuodossa autenttisuuden takaamiseksi.

8 LÄHTEET

Guidelines on the production and preservation of digital audio objects. Second edition. (IASA-TC04). International Association of Sound and Audiovisual Archives, 2009. Elektronista versiota ei ole; tilaustiedot katso http://www.iasa-web.org/technical_guidelines.asp.

Technical guidelines for digital cultural content creation programmes. Version 2.0 September 2008⁴⁶. Minerva project, 2008.

⁴³ <http://www.loc.gov/standards/mets/>

⁴⁴ <http://www.kdk.fi/fi/pitkaaikaissailytyks/maeaerittely-ja-dokumentit/5-suomi/pitkaaikaissailytyks/145-kdkn-hallinnolliset-ja-rakenteelliset-metatiedot-ja-aineiston-paketointi>

⁴⁵ <http://www.kdk.fi/fi/pitkaaikaissailytyks/maeaerittely-ja-dokumentit/5-suomi/pitkaaikaissailytyks/141-kdkn-sailytyks-ja-siirtokelpoiset-tiedostomuodot>

⁴⁶ <http://www.minervaeurope.org/publications/technicalguidelines.htm>

9 TÄRKEIMMÄT ASIASANASTOT

9.1 Kotimaisia

9.1.1 Yleisluontoiset

- Allärs – Allmän tesaurus på svenska; <http://www.kansalliskirjasto.fi/kirjastoala/asiasanastot/allars.html>
- Bella – Specialtesaurus för skönlitteratur; <http://kaunokki.kirjastot.fi/sv-FI/>
- CILLA – Specialtesaurus för musik; <http://www.kansalliskirjasto.fi/kirjastoala/asiasanastot/musacilla.html>
- Kaunokki – Fiktiivisen aineiston asiasanasto; <http://kaunokki.kirjastot.fi/fi-FI/>
- Masa – Museoalan asiasanasto; <http://www.nba.fi/fi/masaetusivu>
- MUSA – Musiikin asiasanasto; <http://www.kansalliskirjasto.fi/kirjastoala/asiasanastot/musacilla.html>
- YSA – Yleinen suomalainen asiasanasto; <http://www.kansalliskirjasto.fi/kirjastoala/asiasanastot/ysa.html>

9.1.2 Tieteenalakohtaiset, laajahko ala

- Agriforest – Maatalous, metsä- ja puuala, elintarvikeala, kotitalous- ja kuluttaja-ala ja ympäristöala; <http://www-db.helsinki.fi/agriforest/>
- Eduskunnan kirjaston asiasanasto – eduskuntatieto, oikeudellinen tieto, yhteiskuntatieto; <http://www.eduskunta.fi/kirjasto/EKS/>
- FinMeSH – Lääketieteen asiasanasto; <http://www.terveysportti.fi/pls/rex/rex.metatesaurus.koti>
- Helecon-asiasanasto – Kauppa- ja talousala; <http://helecon3.hse.fi/helevoc/>
- Kula – Kulttuurien tutkimuksen asiasanasto; <http://www.finlit.fi/kula>
- TEPA-termipankki <http://www.tsk.fi/tepa>

9.1.3 Tieteenalakohtaiset, suppeahko ala

- Hoidokki – Hoitotyön asiasanasto; http://www.shks.fi/hoidokki_hoitotyön_asiasanasto-/
- Kielitieteen asiasanasto; <http://kaino.kotus.fi/www/verkkojulkaisut/asiasanasto.html>
- Kitu – Kirjallisuudentutkimuksen asiasanasto; <http://www.finlit.fi/kitu>
- Käsi- ja taideteollisuuden asiasanasto; <http://www.designkuopio.fi/kirjasto/asiasanasto/>
- Lapponica-asiasanasto; <http://www.lapponica.net/sanasto1.htm>
- Liikunnan ja urheilun asiasanasto; vain painettuna.
- Merenkulun asiasanasto; <http://merisanasto.kyamk.fi/index.php>
- Pankki- ja rahoitussanasto; <http://www.tsk.fi/tsk/pankkisanasto/fi>
- Polas – Poliisin asiasanasto; <http://edui.com:8080/polas/frameset.html>
- Puolustushallinnon asiasanasto; <http://www.mpkk.fi/fi/kirjasto/asiasanasto/>
- Teologisen tiedekunnan kirjaston asiasanasto; <http://www.helsinki.fi/teol/kirjasto/palvelut/asiasanat.htm>
- Tietotekniikan termitalkoot; <http://www.tsk.fi/tsk/termitalkoot/>
- Valokuvan asiasanasto; http://www.fmp.fi/fmp_fi/muvieras/kirjasto/asiasan/index.htm

9.2 Ulkomaisia

- AAT – Art and Architecture Thesaurus; http://www.getty.edu/research/conducting_research/vocabularies/aat/
- ERIC – The Educational Resources Information Center <http://searcheric.org/>
- LCSH – Library of Congress Subject Headings <http://authorities.loc.gov/>
- MeSH – Medical Subject Headings <http://www.nlm.nih.gov/mesh/>
- SOSIG – Social Sciences Information Gateway <http://sosig.esrc.bris.ac.uk/roads/cgi-bin/thesaurus.pl>
- Svenska ämnesord <http://www.kb.se/katalogisering/Svenska-amnesord/>

9.3 Eurooppalaisia

- ETB – European Treasury Browser Project; <http://etb.eun.org/eun.org2/eun/en/etb/content.cfm?lang=en&ov=7208>
- European Education Thesaurus; <http://inet.dpb.dpu.dk/ept/eng/>
- EUROVOC; <http://europa.eu/eurovoc/>
- IATE – Interactive Terminology for Europe, ent. Eurodicautom <http://iate.europa.eu/>
- Iconclass – A classification system designed for art and iconography; <http://www.iconclass.nl/>
- GEMET – General Multilingual Environmental Thesaurus; <http://www.eionet.europa.eu/gemet/about>
- ELSST / HASSET – <http://www.data-archive.ac.uk/search/hassetSearch.asp> (Yhteiskunnallinen tietoaarkisto mukana hankkeessa)

10 STANDARDIEN KUVAUKSET

10.1 Metatietostandardit

Metatieto on aineiston kontekstia, sisältöä ja rakennetta sekä niiden hallintaa ja käsittelyä koko elinkaaren ajan kuvaavaa tietoa. Metatietoa voidaan käyttää muun muassa aineiston hakuun, paikallistamiseen, pitkäaikaissäilyttämiseen ja tunnistamiseen.

Metatiedot jaetaan yleensä kuvailevaan, hallinnolliseen ja rakenteelliseen metatietoon. Kuvaileva metatieto mahdollistaa aineiston löytämisen, tunnistamisen, valinnan sekä käyttöön saamisen ja ymmärrettävyyden. Hallinnollinen metatieto luo edellytykset digitaalisten aineistojen hallinnalle. Sen osa-alueet ovat tekninen, käyttöoikeus- ja pitkäaikaissäilytyksen metatieto. Kuvaileva ja hallinnollinen metatieto eivät ole tiukasti erotettavissa toisistaan; käytännössä kuvaileva metatieto sisältää joitakin hallinnollisen metatiedon elementtejä, mutta ei kattavasti.

Metatiedot tuotetaan usein kuvailusääntöjen perusteella. Nämä säännöt, kuten kirjastojen anglo-amerikkalaiset luettelointisäännöt, ovat olleet sektorikohtaisia, toisin sanoen kirjastojen säännöt on kehitetty vain kirjastoja ajatellen, ja niiden sovellettavuus arkisto- ja museoaineistolle on ollut rajallinen. Kuvailuperiaatteiden yhtenäistämiseksi Kansallinen digitaalinen kirjasto -hankkeessa pyritään soveltamaan uusia Resource Description and Access – eli RDA-sääntöjä, jotka on kehitetty kirjastosektorilla mutta jotka periaatteessa soveltuvat myös muille sektoreille.

Kuvailevan metatiedon pohjana olevat formaatit ovat sektorikohtaisia. Kirjastot soveltavat MARC-formaattia, arkistot EAD:tä ja museot LIDOa. Kaksi viimeksi mainittua ovat verraten uusia ratkaisuja, ja varsinkin

museoissa tallennus tapahtuu yleensä jossakin muussa muodossa, josta metatieto muunnetaan siirtoa varten LIDOon. Kirjastoissa MARC on myös tallennusformaatti, ja sen soveltamisen periaatteet ovat hyvin yhtenevät eri kirjastoissa. Arkisto- ja museoaineiston kuvailussa saattaa olla enemmän organisaatiokohtaisia eroja.

Aineistoja koskevien metatietojen ohella on tarpeen siirtää myös henkilöitä ja organisaatioita koskevia metatietoja eli ns. auktoriteettitietoja. Niiden ylläpidossa tehdään paljon päällekkäistyötä, ja nimenmuotojen erot (esim. Axel Gallen – Akseli Gallen-Kallela) vaikeuttavat tiedonhakuja. Auktoriteettitietojen vaihdossa käytettävät tietoelementit ja syntaksi tulee sopia erikseen. Sektoreiden tulisi asettaa auktoriteettitietonsa tarjolle erillisiin sektorikohtaisiin auktoriteettitietokantoihin, jotka tulisi synkronoida keskenään. Tämä menettely helpottaisi myös esimerkiksi nimien kansainvälisen standarditunnisteen (ISNI) käyttöönottoa Suomessa.

Hallinnollisessa metatiedossa sektorikohtaisia eroja ei a priori ole, koska esimerkiksi still-kuvan tekniset ominaisuudet ovat aina samat, riippumatta siitä minkä tyyppinen organisaatio tiedostoa hallinnoi.

10.2 Rajapintastandardit

10.2.1 Metatietojen haravointi

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) määrittelee rajapinnan kuvailevan metatiedon haravointiin taustajärjestelmistä. Haravointi voidaan tehdä valikoiden (halutulta aikaväliltä; vain uudet / muokatut / poistetut tietueet). Metatieto voidaan siirtää yleisesti tunnetuissa formaateissa kuten Dublin Core tai MARC-muodossa.

The Open Archives Initiative ja ANSI/NISO kehittävät OAI-PMH:n pohjalta uutta standardia työnimeltään ResourceSync. Sen tavoitteena on luoda edellytykset sekä metatiedon että digitaalisten objektien haravoimiselle⁴⁷.

10.2.2 Haku

Search/Retrieve via URL (SRU) on Kongressin kirjaston ylläpitämä yleinen tiedonhakustandardi. Sen edeltäjä, ANSI/NISO Z39.50 Information retrieval application service definition and protocol specification (julkaistu ISO-standardina ISO 23950), on edelleen laajalti käytössä kirjastojärjestelmissä.

Tiedonhakustandardit määrittelevät laitteisto- ja ohjelmistoriippumattoman tavan hakukyselyiden ja hakutulosten siirtämiseen. Tekniset toteutukset vaihtelevat: SRU on Web-ympäristöön sopiva yhteyden protokolla joka perustuu http:n käyttöön, kun taas Z39.50 on yhteydellinen ja soveltaa ISO:n OSI-malliin perustuvia tietorakenteita. SRU:ta ja Z39.50:ttä yhdistää mahdollisuus määrittellä eksaktisti haussa käytettävät termit (kuten tekijä, nimeke, ISBN-tunniste, asiasana jne.).

Standardeihin perustuvia avoimia hakurajapintoja on toteutettu lähinnä vain kirjastojärjestelmiin. Rajapintojen puuttuminen aiheuttaa runsaasti ongelmia: jos haku perustuu tietokannan normaalin asiakasliittymän antaman palautteen analysointiin, pienetkin muutokset esim. asiakasliittymän ulkoasussa voivat haitata hakuja tai estää sen kokonaan. Silloinkin kun on olemassa standardirajapinta tiedonhaku voi olla vaikeaa kohdetietokantojen semanttisten erojen vuoksi: tarvittava hakutermejä ei ole lainkaan, tai sitä ei ole indeksoitu.

⁴⁷ <http://www.niso.org/workrooms/resourcesync/>

Z39.50:ttä ja SRU:ta on pidetty kirjastojen standardeina. OpenSearch⁴⁸ on Amazonin kehittämä Web-pohjainen avoin hakurajapinta, josta pyritään tekemään IETF-standardi. OpenSearch ei edeltäjiensä tavoin näytä pystyneen mainittavaan läpimurtoon. Googlen Search Appliance⁴⁹ voi osoittautua edeltäjiään menestyksekkäämmäksi. KDK:lle se ei välttämättä sovellu, koska lähtökohtana ei ole metatiedon vaan kokotekstin haravointi.

10.2.3 Sovellustason rajapinnat

SRU:n ja OAI-PMH:n kaltaiset rajapinnat tarjoavat mahdollisuuden tiedonhakuun ja/tai metatietojen haravointiin. Käytettävien sovellusten määrän kasvaessa tarvitaan monipuolisempia standardeja, jotka mahdollistavat periaatteessa minkä tahansa johonkin järjestelmään tallennetun metatiedon hyväksikäytön muissa sovelluksissa. Digital Library Foundation on kehittänyt Integroitujen kirjastojärjestelmien sisältämien tietojen (esimerkiksi asiakastiedot, lainaustiedot jne.) siirtoon DLF ILS Discovery Interface Task Group (ILS-DI) Technical Recommendation -suosituksen. Kirjastojärjestelmille on kehitetty myös kapea-alaisempia rajapintamäärittäjiä, kuten ANSI/NISO:n Circulation Interchange Protocol, joka nimensä mukaisesti keskittyy lainaustietoihin.

10.3 Tunnisteet

Asiakasliittymän ja PAS-ratkaisun tietosisältöjen haku- ja päivitysoperaatioiden on kohdistuttava oikein. Tämän varmistamiseksi jokaisella asiakasliittymässä ja PAS-ratkaisussa olevalla tietosisällöllä on oltava pysyvä, ainutkertainen ja toiminnallinen tunnus, jota voidaan käyttää tiedonhaussa ja josta voidaan muodostaa pysyvä web-osoite. Tietosisältöjä voivat olla esimerkiksi julkaisut, asiakirjat, museoesineet, mutta myös esimerkiksi metatietoviitteet tai useampia dokumentteja sisältävät paketit.

Jos tietosisältö on rakenteinen (esimerkiksi digitoitu kirja, jonka jokainen sivu on erillinen kuva, tai XML -muodossa oleva artikkeli johon liittyy kuvia), yksittäisille tiedostoille on annettava oma tunnus, koska tiedostoja voidaan myöhemmin migroida toiseen tiedostomuotoon. Jos migraatiotarvetta ei ole (tietosisältö on esimerkiksi ohjelma, jota ei voida muuntaa) on suositeltavaa antaa tunniste, etenkin jos tietolähteeseen halutaan luoda suora Internet-linkki.

Tunnisteiden jakelua vaikeuttaa se, että kaikille aineistoille ei vielä ole standarditunnisteita. Still-kuvat ovat esimerkki tästä. Toisaalta muutamia valmiitakaan tunnistejärjestelmiä kuten tekstimuotoisten teosten ISTC:tä tai henkilö- ja yhteisötekijöiden ISNI:ä ei vielä ole otettu Suomessa käyttöön. Jos standarditunniste puuttuu, voidaan soveltaa Kansalliskirjaston NBN (National Bibliography Number) -tunnistetta, jonka jakelu ja resoluutio URN-tunnisteena ovat hajautettavissa. Mutta olemassa olevan tunnisteiden kuten ISTC:n korvaaminen NBN:llä on vaikeata, koska tällöin Suomi jäisi pysyvästi kansainvälisen yhteistyön ulkopuolelle. Tästä syystä KDK:n pitäisi tukea esimerkiksi teosten tunnisteiden käyttöönottoa Suomessa.

Suomessa käytössä olevilla standarditunnisteilla on yleensä tukiorganisaatio, kuten kansallinen ISBN-keskus⁵⁰, jolta voi pyytää apua tunnisteiden soveltamista koskeissa ongelmatilanteissa.

⁴⁸ <http://www.opensearch.org>

⁴⁹ <http://code.google.com/apis/searchappliance/documentation/614/>

⁵⁰ <http://www.kansalliskirjasto.fi/julkaisuala/isbn.html>

10.4 Pitkäaikaissäilytys ja aineistojen siirto

Pitkäaikaissäilytyksen standardoinnilla pyritään siihen, että digitaaliset aineistot voidaan säilyttää käyttökuntoisina useiden vuosikymmenien ajan erilaisten säilytysmenetelmien avulla ja mahdollistetaan tehokas aineistojen siirto taustajärjestelmien PAS-ratkaisun välillä.

Yleisesti käytetty OAIS-viitemalli määrittelee pitkäaikaissäilytyksen yleisen viitekehyksen, keskeiset käsitteet, toimijat ja prosessit. Viitemallia sovelletaan käytännössä kaikissa digitaalisen pitkäaikaissäilytyksen hankkeissa. Sen luoma yhtenäinen käsitteistö on suuresti helpottanut pitkäaikaissäilytystä koskevaa tieteellistä keskustelua, käytännön toimintaa ja kehitystä.

OAIS-viitemallin ohella pitkäaikaissäilytyksen keskeisiä standardeja ovat container- eli säiliöstandardit. OAIS määrittelee, että PAS-toimintaan liittyy erilaisia paketteja, kuten pitkäaikaissäilytysorganisaatiolle lähetettävä siirtopaketti. Lisäksi malli kertoo, millaisiin osiin nämä paketit jakautuvat. Mutta OAIS ei pyri määrittelemään täsmällisesti tämän paketin rakennetta. Tästä vastaavat säiliöstandardit, joita KDK:ssa on kolme: yleiskäyttöinen METS, verkkoarkistojen WARC sekä liikkuvan kuvan MXF. Näiden standardien avulla voidaan paketoita yhteen sekä itse dokumentti kaikkine tiedostoineen, että dokumenttiin liittyvät hallinnolliset ja kuvailevat metatiedot. Nämä standardit vastaavat tavaraliikenteen kontteja siinä, että ne helpottavat digitaalisten aineistojen käsittelyä.

METS-standardin valintaa tulee täydentää ohjeistuksella sen soveltamisesta. Kansallinen digitaalinen kirjasto -hanke ylläpitää erillisistä määrittelyjen aineistojen paketoinnista⁵¹.

Pitkäaikaissäilytys edellyttää korkeatasoisia hallinnollisia metatietoja. Standardisalkku listaa käytettävät metatietoformaattit (kuten still-kuvien teknisen metatiedon MIXin) mutta näitä metatietoja sekä dokumenttien tiedostomuotoja ei esitellä standardisalkussa, vaan sitä tukevassa muussa ohjeistuksessa.

Pitkäaikaissäilytystä koskeva ohjeistus ja määritykset täsmentyvät ennen varsinaisen PAS-toiminnan alkua. Osa KDK:n PAS-ratkaisun toimintaa on näiden KDK:n kokonaisarkkitehtuurissa standardisalkkuun rinnastettavien määritysten ylläpitäminen.

⁵¹ <http://www.kdk.fi/fi/pitkaaikaissailytys/maeaerittely-ja-dokumentit/5-suomi/pitkaaikaissailytys/145-kdkn-hallinnolliset-ja-rakenteelliset-metatiedot-ja-aineiston-paketointi>



Kansallinen
digitaalinen
kirjasto