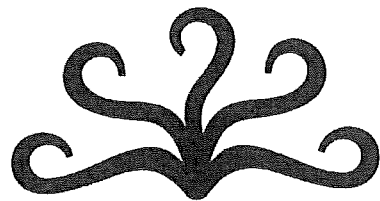


KANSALLISEN DIGITAALISEN KIRJASTON STANDARDISALKKU

v1.1.0



Kansallinen
digitaalinen
kirjasto

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



Lisenssi

Creative Commons Suomi CC-BY-NC-SA 4.0 (<http://creativecommons.org/licenses/by-nc-sa/4.0/deed.fi>)

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

- *Opetus- ja kulttuuriministeriö nimetään teoksen tekijäksi (ei kuitenkaan siten, että ilmoitus viittaisi lisenansiantajan tukevan lisenssinsajaaja tai teoksen käyttötapaaj).*
- *Epäkaupallinen osapuoli eli käyttäjä ei saa käyttää teosta kaupallisesti.*
- *Mikälii 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	7
3.1	Kuvailusäännöt.....	7
3.2	Kuvailevan metatiedon formaatit.....	8
3.3	Henkilöiden, sukujen ja yhteisöjen tiedot.....	9
3.4	Sisällönkuvailuun käytettävät välineet.....	9
3.5	Hallinnollinen metatieto.....	9
3.6	Rakenteellinen metatieto.....	10
4	RAJAPINNAT	11
4.1	Metatiedon haravointi.....	11
4.2	Haku.....	11
4.3	Sovellustason rajapinnat.....	12
5	LINKITYSPALVELUT JA TUNNISTEET	13
5.1	Linkitys.....	13
5.2	Tunnisteet: aineisto.....	13
5.3	Tunnisteet: tekijät.....	13
6	AUTENTIKOINTI JA AUKTORISOINTI	14
6.1	Tiivistealgoritmit.....	14
7	PITKÄAIKAISSÄILYTYS JA AINEISTOJEN SIIRTO PAS-PALVELUUN	15
8	TÄRKEIMMÄT ASIASANASTOT	16
8.1	Kotimaisia.....	16
8.1.1	Yleisluontoiset.....	16
8.1.2	Tieteenalakohtaiset, laajahko ala.....	16
8.1.3	Tieteenalakohtaiset, suppeahko ala.....	16
8.2	Ulkomaisia.....	17
8.3	Eurooppalaisia.....	17
9	STANDARDIEN KUVAUKSET	18
9.1	Metatietostandardit.....	18
9.2	Rajapintastandardit.....	18
9.2.1	Metatietojen haravointi.....	18
9.2.2	Haku.....	19
9.2.3	Sovellustason rajapinnat.....	19
9.3	Tunnukset.....	19
9.4	Pitkäaikaissäilytys ja aineistojen siirto.....	20



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 alisalkkuineen (katso <https://www.avoindata.fi/> (haku standardisalkku) sekä <http://www.jhs-suositukset.fi/suomi/jhs181>).



1 JOHDANTO

Kansallisen digitaalisen kirjaston (KDK) 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 sovellosohjeeseen (profiliin), mutta myös kansallista ohjeistusta tarvitaan. KDK:n oman salkun ohella hankkeessa sovelletaan myös Julkishallinnon standardisalkkua alisalkkuineen¹.

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

Standardisalkussa määritellyt 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.

KDK:n standardisalkku sisältää sekä virallisia että ns. teollisuusstandardeja. Digitaalisen kirjaston standardipohja on yhä osittain rakenteilla, joten salkkuun on päivitysten yhteydessä lisätty uusia standardeja tai päivitetty jo salkussa olevien standardien versioita. Kaikki listatut standardit ovat avoimia eli sekä standardit itse että niitä koskeva dokumentaatio on julkisesti käytettävissä². Osa salkun standardeista on maksuttomia, osa maksullisia.

Standardisalkkua on täydennetty seuraavasti:

1. Käyttöoikeuksien metatiedosta sekä eri aineistotyyppien (teksti, audio, still-kuva, elävä kuva) teknisestä metatiedosta on laadittu erilliset suositukset.
2. Yksittäisille standardeille on tarvittaessa tehty soveltamisohjeita eli profiileja ohjelmistokehittäjien ja järjestelmien ylläpitäjien tueksi. Esimerkki tällaisesta profiilista on ohje aineistojen ja niiden metatietojen paketoinnista pitkäaikaissäilytykseen, joka ohjaa METS- ja PREMIS-standardien soveltamista.

Salkun ylläpidosta vastaa KDK-hankkeen ohjausryhmä.

¹ <https://www.avoindata.fi>, <http://www.ihs-suositukset.fi/suomi/ihs181>

² Standardien avoimuudesta katso https://en.wikipedia.org/wiki/Open_standard



2 YLEISET STANDARDIT

Asiakasliittymän ja PAS-palvelun on tuettava UNICODE-standardin³ versiota 7.X tai uudempaa (UTF-8 -koodauksella) **(1)**.

Viitetieto ja tekstidokumentit on suositeltavaa toimittaa taustajärjestelmistä UTF-8 -koodattuina. **(2)**. Yksittäisten tietoelementtien merkkivalikoimasta voi olla olemassa tarkempia ohjeita. Esimerkiksi useiden tunnusjärjestelmien merkkivalikoima on tarkoin rajattu, ja silloinkin kun se ei ole – esimerkkinä Handle system ja DOI – on suositeltavaa käyttää tunnuksessa vain tulostettavia ISO Latin 1 -merkkejä. **(2)**

Asiakasliittymän ja PAS-palvelun käyttöliittymien on tuettava Web Content Accessibility Guidelines (WCAG) -suosituksen versiota 2.0⁴ vähintään tason A vaatimusten mukaisesti⁵ **(1)**, suositeltavan tason ollessa AA **(2)**. Käyttöliittymien suunnittelussa ja järjestelmien asiakkaille suunnattujen ohjeiden julkaisemisessa on suositeltavaa ottaa huomioon julkishallinnon Esteettömyyssalkun⁶ vaatimukset.

³ <http://unicode.org/>

⁴ <http://www.w3.org/TR/WCAG20>

⁵ <http://www.w3.org/WAI/WCAG20/quickref/>

⁶ <https://www.avoindata.fi/data/fi/dataset/esteettomyyssalkku>



3 METATIEDOT

KDK:n järjestelmäkokonaisuuden toiminta edellyttää riittäviä ja sisällöllisesti yhteneviä metatietoja, joten metatietoformaateista ja kuvailun periaatteista sopiminen on KDK:n kannalta välttämätöntä. Nykytilanne on haasteellinen sen vuoksi, että museosektorilla ei ole ollut yhteisiä kuvailusääntöjä, ja arkistosektorillakin erityisarkistoilla on ollut omat kuvailuperiaatteensa. Vain kirjastosektori on ollut tässä suhteessa yhtenäinen, mutta kirjastojenkin kuvailusäännöt (ja metatietoformaatti) ovat uusimisen tarpeessa.

Metatieto voidaan jakaa perinteiseen kuvailevaan sekä hallinnolliseen metatietoon. Erityyppisille metatiedoille sovelletaan eri formaatteja, joiden tietoelementit ovat osittain päällekkäisiä⁷.

Viranomaisten asiakirja-aineistojen on noudatettava arkistolaitoksen Sähke-määräysten⁸ linjauksia **(1)**.

3.1 Kuvailusäännöt

Muistiorganisaatiot tekevät kansainvälistä yhteistyötä eri sektorien käsitelmien harmonisoimiseksi. Kansallisesti yhteentoimivuutta pyritään edistämään mm. yhteisten ontologioiden käyttöönotolla ja yhteistä nimitietopalvelua suunnitellaan.

KDK-hankkeessa kuvailun kehittämisen perustaksi suositellaan kirjastojen RDA-säännöstöä (Resource Description and Access⁹) **(2)**, vaikka sektoreilla tai yksittäisillä organisaatioilla voi lisäksi olla käytössä myös omia kuvailusäännöstöjä.

RDA tukee kirjastosektorin aiempia kuvailusääntöjä paremmin elektronisten aineistojen kuvailua sekä teosten ja niiden manifestaatioiden linkittämistä toisiinsa. RDA on julkaistu 2011, ja ensimmäiset kirjastot ottivat sen käyttöön jo vuonna 2013. Arkistoissa ja museoissa sitä ei tiettävästi vielä käytetä. Suomalainen versio julkaistiin 2015, ja sitä alettiin soveltaa kirjastosektorilla vuoden 2016 alussa. Käännöstyöstä ja RDA:n käyttöönotosta Suomessa vastaa Kansalliskirjasto. Kansallisen sovelluksen teossa on pyritty ottamaan KDK:n ja eri sektoreiden tarpeet huomioon.

Arkistosektorilla on kansainvälisesti käytössä International Council of Archives (ICA) –järjestön puitteissa tuotetut yleiset kuvailustandardit:

- *International Standard Archival Description* (ISAD(G) 1997),
- *International Standard Archival Authority Record for Corporate Bodies, Persons, and Families* (ISAAR(CPF) 2004) ja
- *International Standard for Describing Functions* (ISDF 2008).

Lisäksi ICA:n EGAD-työryhmä valmistelee arkistokuvailun kansainvälistä käsitemallia RiC "Records in Contexts" ja sitä tarkentavaa ontologiaa.

Erityisarkistoilla on omat kansainvälisesti yhdenmukaiset käytäntönsä (esim. data-arkistot ja elokuva-arkistot). Kansallisella tasolla on julkaistu luonnos em. standardeihin sekä RDA:han ja ISO 21127 A *Reference ontology for the interchange of cultural heritage information* (2006, uusin versio 2014) standardiin perustuva arkistokuvailun uusi kansallinen käsitemalli, jonka tavoitteena on toimia erilaisten metatietomallien metamallina. Käsitemalliin pohjautuvien uusien yleisten kansallisten arkistokuvailusääntöjen laadintaan ryhdytään sen jälkeen, kun ICA:n kansainvälisten kuvailustandardien kehittäminen valmistuu. Nämä säännöt

⁷ <http://www.loc.gov/marc/bibliographic/mapping/index.html>

⁸ <http://www.arkisto.fi/fi/palvelut/julkisen-hallinnon-saehkoeiset-palvelut/saehke-maeaeraeykset>

⁹ <http://rdatoolkit.org/>



tulevat korvaamaan vuonna 1997 julkaistut *Arkistojen kuvailu- ja luettelointisäännöt*. Ahaa -hankkeen puitteissa tehdään järjestelmän kuvailuohje, jota voidaan hyödyntää kansallisten sääntöjen tuottamisvaiheessa.

Museosektorilla on vuodesta 2014 alkaen sovellettu

- Iso-Britanniassa Collections Trust -säätiön laatimaa ja ylläpitämää SPECTRUM-luettelointistandardia¹⁰.

Sen kotimainen versio on Museo 2015-hankkeessa¹¹ laadittu Museoiden luettelointiohje¹², jota suositellaan käytettävän museaalisten kokoelmien kuvailussa ja luetteloinnissa (2).

3.2 Kuvailevan metatiedon formaatit

Kukin sektori voi käyttää kuvailussa yhtä tai useampaa metatietoformaattia. Mahdolliset sovelluskohtaiset formaatit tulee muuntaa ennen asiakasliittymään tai PAS-palveluun siirtämistä johonkin yleisesti käytettyyn vaihtoformaattiin. KDK:n yhteisissä tietojärjestelmissä metatiedot tallennetaan ilman konversioita siinä muodossa, jossa ne on vastaanotettu. Kunkin formaatin tietoelementtien indeksoinnista päätetään Finnassa erikseen. Tämä onnistuu helpoimmin jos kuvailuperiaatteet ja formaattien soveltamistavat ovat riittävän yhtenäiset.

PAS-palvelussa kuvailevaa metatietoa tarvitaan pitkäaikaissäilytyksen tueksi. Näiden metatietojen mahdollisesta migraatiosta sovitaan tietojen tuottajan ja PAS-palvelun kesken erikseen.

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, tai kuvailuformaatteihin pohjautuvia siirtoformaatteja, 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.

¹⁰ <http://www.collectionstrust.org.uk/spectrum/the-spectrum-standard>

¹¹ http://www.nba.fi/fi/museoalan_kehittaminen/museo_2015/museoiden_luettelointiohje

¹² <http://www.luettelointiohje.fi/>



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 kunkin yhteisen elementin sisällön muodostamisen osalta.

3.3 Henkilöiden, sukujen ja yhteisöjen tiedot

Henkilöt, suvut ja yhteisöt kuvaillaan kirjasto-, arkisto- ja museosektoreilla RDA-sääntöjen mukaisesti. Metatiedot suositellaan tallennettavaksi niin kattavina, että ne mahdollistavat yksiselitteisen tunnistamisen **(2)**. Mahdollisen yhteisen nimitietopalvelun formaatti sovitaan erikseen. Sektorit käyttävät nimitietojen tallentamiseen joko yllä mainittuja kuvailevan metatiedon formaatteja tai sektorikohtaisia ns. auktoriteettiformaatteja, joita ovat:

- MARC 21 Auktoriteettitiedot
- EAC-CPF (Encoded Archival Context – Corporate bodies, Persons and Families)

Konversiot näistä formaateista nimitietopalvelussa sovellettavaan formaattiin määritellään erikseen.

Poimittaessa nimitietoja muista julkishallinnon järjestelmistä kuten Väestötietojärjestelmästä tai Patentti- ja rekisterihallituksen rekistereistä vaihtoformaattina pyritään soveltamaan EU:n Core Person - ja Core Business -määrittelyjä¹⁴.

ISNI-tietokannan rajapinnat määritellään erikseen.

3.4 Sisällönkuvailuun käytettävät välineet

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

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

Asiasanastojen ja ontologioiden lisäksi sisällönkuvailuun voidaan käyttää kansainvälisiä luokitusjärjestelmiä, kuten Outline of Cultural Materials (OCM), Iconclass ja Classification of Individual Consumption according to Purpose (COICOP).

3.5 Hallinnollinen metatieto

Kuvailusäännöt ja kuvailevan metatiedon formaatit eivät kata kaikkea hallinnollista metatietoa. Sen vuoksi KDK:ssa on laadittu erilliset ohjeet hallinnollisesta (ja rakenteellisesta) metatiedosta¹⁵. Kaikelle pitkäaikaissäilytettävälle aineistolle on luotava KDK-ohjeistuksen mukaiset tekniset metatiedot **(1)**.

Hallinnollisen metatiedon osalta kaikilla sektoreilla on päästävä riittävän yhteismitallisiin ratkaisuihin. Kyseessä on uusi toiminta-alue, jolla ei ole vakiintuneita käytänteitä, minkä vuoksi riittävän ohjeistuksen laadinta ja PAS-palveluun vietävien siirtopakettien testaaminen ovat tärkeitä.

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

¹⁴ https://joinup.ec.europa.eu/asset/core_vocabularies/description

¹⁵ <http://www.kdk.fi/fi/pitkaaikaissailytys/maaeerittely-ja-dokumentit>



Teknisen metatiedon formaatteja on käytettävissä vain yksi kutakin aineistotyyppiä kohden.

Tekninen metatieto

Vaaditaan seuraavien formaattien käyttöä soveltuvin osin Säilytys- ja siirtokelpoiset tiedostomuodot määrittämisen mukaisesti **(1)**:

- MIX (NISO Metadata for images in XML Schema)¹⁶
- ADDML (Archival Data Description Markup Language)¹⁷
- AudioMD (Technical metadata for Audio)¹⁸
- VideoMD (Technical metadata for Video)

Pitkäaikaissäilytyksen metatieto

Seuraavien formaattien käyttö on välttämätöntä **(1)**:

- PREMIS Data dictionary for preservation metadata, versio 2.2.

Käyttöoikeustiedot

PAS-palvelua koskevat käyttöoikeustiedot – kuten sen kuvaaminen, keillä on oikeus tehdä migraatio arkistoidulle dokumentille - tallennetaan PREMIS-metatietoformaatin¹⁹ Rights-osioon erikseen sovittavalla tavalla **(1)**.

Taustajärjestelmiin suositellaan tallennettavaksi digitaalisen aineiston tekijänoikeudellinen status sekä – jos aineistolla on tekijänoikeussuoja – mahdollinen käyttöä koskeva lisenssi. Lisäksi on tarvittaessa kuvattava mahdolliset lakeihin tai sopimuksiin perustuvat käyttörajoitukset **(2)**.

KDK:ssa on laadittu yleisohje käyttöoikeusmetatietojen tallennuksesta, mutta sen lisäksi sektorien pitää laatia omat, formaattikohtaiset ohjeet. Tätä kirjoitettaessa (maaliskuu 2016) kirjastojen ohje on jo valmis ja sitä sovelletaan käytännössä. Sektoritason ohjeistus luo edellytykset sille, että käyttöoikeustiedot ovat siirrettävissä Finnaan KDK:n tukemassa vaihtoformaattissa ja sovellettavissa siellä.

Lisenssin ja muiden sopimusten sisältöä ei yleensä tarvitse kuvata metatiedoissa; riittää että mainitaan lisenssin nimi ja tarjotaan linkki lisenssin verkossa olevaan kuvaukseen. Käyttörajoitusten (aineistot ja niiden metatiedot) kuvaamiseen tarvitaan peruste (laki tai sopimus), lain / sopimuksen nimi, käyttörajoituksen selite sekä rajoituksen kesto.

3.6 Rakenteellinen metatieto

PAS-järjestelmään siirrettävien dokumenttien rakenteen kuvaamiseen on käytettävä METS StructMap -standardia **(1)**.

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

¹⁷ <http://www.arkivverket.no/arkivverket/Arkivbevaring/Elektronisk-arkivmateriale/Standarder/ADDML>

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

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



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.

Metatietojen haravointiin sovelletaan OAI-PMH -protokollan (The Open Archives Initiative Protocol for Metadata Harvesting) versiota 2.0²⁰ **(1)** tai vaihtoehtoisesti Atom-standardia^{21,22} **(3)**.

OAI-PMH:n toteutuksessa on suositeltavaa noudattaa ohjetta Best practices for OAI-PMH DataProvider implementations and shareable metadata²³ **(2)**.

OAI-PMH:lla haravoitavien aineistojen tulisi tarjota vähintään OAI-PMH DC XML -skeeman²⁴ mukaisia tietueita **(2)**. Muiden skeemojen kuten MARC-formaatin käytöstä sekä mahdollisista laajennuksista tai muutoksista DC XML -skeemaan sovitaan erikseen.

Atom-protokollan käytöstä sovitaan erikseen.

Metatietojen indeksointijärjestelyistä asiakasliittymässä ja mahdollisista normalisointijärjestelyistä Formula-formaattimuuntimessa sovitaan erikseen.

4.2 Haku

Asiakasliittymässä on oltava SRU-palvelin, joka tukee

- SRU-standardin versiota 1.2²⁵ **(1)** ja
- SRU:n palvelua Search/Retrieve Operation **(1)**.

Asiakasliittymässä on suositeltavaa olla OpenSearch-hakurajapinta²⁶ **(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) -standardin²⁷ tuki **(2)**. Elleivät nämä standardit ole käytettävissä, sovellusten pitäisi tukea Simple Object Access Protocol (SOAP) -standardin versiota 1.2²⁸ tai Representational State Transfer (REST) -määrittystä²⁹ **(2)**.

²⁰ <http://www.openarchives.org/OAI/openarchivesprotocol.html>

²¹ <http://tools.ietf.org/html/rfc4287>

²² <http://bitworking.org/projects/atom/rfc5023.html>

²³ <http://www.diglib.org/pubs/dlf108.pdf>

²⁴ http://www.openarchives.org/OAI/2.0/oai_dc.xsd

²⁵ <http://www.loc.gov/standards/sru/index.html>

²⁶ <http://www.opensearch.org/Home>

²⁷ <http://www.niso.org/workrooms/ncip>

²⁸ <http://www.w3.org/TR/soap/>

²⁹ 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 palvelurajapinta³⁰. Muut taustajärjestelmät voivat **(3)** tukea ILS-DI:tä.

³⁰ http://old.diglib.org/architectures/ilsdi/DLF_ILS_Discovery_1.0.pdf



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 OpenURL³¹-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 standarditunnisteita, 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 asiakasliittymässä että PAS-palvelussa (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 159³².

5.3 Tunnisteet: tekijät

Henkilö- ja yhteisötekijöiden tunnistamiseen on suositeltavaa käyttää International Standard Name Identifier (ISNI) –standarditunnusta (ISO 27729)³³ **(2)**. Tutkijoiden tunnistamiseen voidaan ISNI:n rinnalla käyttää ORCID-tunnusta³⁴ **(3)**, joka vastaa rakenteeltaan ISNIä, mutta jolla on eri ylläpito-organisaatio ja jossa vastuu tunnisteiden luonnista on hajautettu tutkijoille, kun ISNI-järjestelmässä vastuu on keskitetty kansainväliselle ISNI-keskukselle.

KDK:n yhteisen nimitietokannan sekä sektorikohtaisten järjestelmien kuten kirjastojen Asterin on tuettava sekä ORCID- että ISNI-tunnuksia **(1)**.

³¹ <http://en.wikipedia.org/wiki/OpenURL>

³² <http://www.jhs-suositukset.fi/suomi/jhs159>

³³ <http://isni.org/>

³⁴ <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 7.8.2009/617³⁵ 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 tai SSH-protokollan³⁶ 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 asiakasliittymässä 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)**.

³⁵ <http://www.finlex.fi/fi/laki/ajantasa/2009/20090617>

³⁶ <https://tools.ietf.org/html/rfc4252>

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 PAS- PALVELUUN

Pitkäaikaissäilytyksessä tulee soveltaa CCSDS:n (The Consultative Committee for Space Data Systems) julkaisemaa OAIS-viitemallia⁴³ (Open Archival Information System), joka on julkaistu myös ISO-standardina ISO 14721:2012 **(1)**. Standardista on julkaistu myös suomenkielinen versio ”Viitemalli pitkäaikaissäilytysarkistolle” (SFS 5972 2015).

OAIS-viitemallissa pitkäaikaissäilytykseen siirrettävää yksittäistä aineistokokonaisuutta kutsutaan siirtopaketti. Paketti sisältää sekä itse aineiston että sitä kuvaavan hallinnollisen ja kuvailevan metatiedon. ”Paketointiin” tulee soveltaa METS (Metadata Encoding and Transmission Standard) -standardin⁴⁴ versiota 1.11 **(1)**, ja METS-pakettien rakenteen tulee olla KDK:n paketointimäärityksen mukainen⁴⁵ **(1)**.

PAS-palveluun siirrettävän aineiston tulee olla Kansallisen digitaalisen kirjaston sallimissa säilytys- tai siirtokelpoisissa tiedostomuodoissa⁴⁶ **(1)**, mutta dokumentit voidaan sisällyttää pakettiin myös alkuperäisessä tiedostomuodossa autenttisuuden takaamiseksi. Audiotallenteille voidaan lisäksi soveltaa IASA:n 2009 julkaisemia periaatteita [IASA] ao. tiedostomuotojen osalta⁴⁷ **(2)**. Tiedostoihin on liitettävä Kansallisen digitaalisen kirjaston ohjeistuksen mukainen hallinnollinen, rakenteellinen ja kuvaileva metatieto⁴⁸.

Sovellustason rajapintojen toteuttamiseen käytetään SSH File Transfer Protocol (Secure FTP, SFTP) -standardia⁴⁹, HTTPS-tiedonsiirto-protokollaa⁵⁰ ja REST-rajapintastandardia⁵¹. Rajapintoihin sisältyvää toiminnallisuutta kehitetään yhteistyössä hyödyntävien organisaatioiden kanssa.

PAS-palvelun auditoinnissa sovellettava standardi päätetään myöhemmin. Teknisesti kattavin vaihtoehto on CCSDS:n OAIS-malliin pohjautuva Audit and certification of trustworthy digital repositories⁵², joka on julkaistu ISO-standardina ISO 16363. Koska siihen perustuvaa sertifiointia ei toistaiseksi ole mahdollista hankkia, sovelletaan muita OAIS-pohjaisia vaihtoehtoja kuten Data seal of approval -määritystä⁵³ tai World Data System’in WDS-sertifikaattia⁵⁴.

⁴³ <http://public.ccsds.org/publications/archive/650x0m2.pdf>

⁴⁴ <http://www.loc.gov/standards/mets/>

⁴⁵ <http://www.kdk.fi/fi/pitkaaikaissailytyks/maeaerittely-ja-dokumentit/5-suomi/pitkaaikaissaailytyks/145-kdkn-hallinnolliset-ja-rakenteelliset-metatiedot-ja-aineiston-paketointi>

⁴⁶ <http://www.kdk.fi/fi/pitkaaikaissailytyks/maeaerittely-ja-dokumentit/5-suomi/pitkaaikaissaailytyks/141-kdkn-sailytyks-ja-siirtokelpoiset-tiedostomuodot>

⁴⁷ <http://www.iasa-web.org/tc04/audio-preservation>

⁴⁸ <http://www.kdk.fi/fi/pitkaaikaissailytyks/maeaerittely-ja-dokumentit/5-suomi/pitkaaikaissaailytyks/145-kdkn-hallinnolliset-ja-rakenteelliset-metatiedot-ja-aineiston-paketointi>

⁴⁹ <https://tools.ietf.org/html/draft-ietf-secsh-filexfer-13>

⁵⁰ <https://tools.ietf.org/html/rfc7230>

⁵¹ <https://tools.ietf.org/html/rfc7231>

⁵² <http://public.ccsds.org/publications/archive/652x0m1.pdf>

⁵³ <http://datasealofapproval.org/en/>

⁵⁴ <https://www.icsu-wds.org/community/membership/certification>



8 TÄRKEIMMÄT ASIASANASTOT

8.1 Kotimaisia

8.1.1 Yleisluontoiset

- Allärs – Allmän tesaurus på svenska; <http://finto.fi/allars/sv/>
- CILLA – Specialtesaurus för musik; <http://finto.fi/musa/sv/>
- JUHO – Julkishallinnon ontologia; <http://finto.fi/juho/fi/>
- Kauno – Fiktiivisen aineiston ontologia; <http://finto.fi/kauno/fi/>
- Masa – Museoalan ja taideteollisuusalan ontologia; <http://finto.fi/maotao/fi/>
- MUSA – Musiikin asiasanasto; <http://finto.fi/musa/fi/>
- YSA – Yleinen suomalainen asiasanasto; <http://finto.fi/ysa/fi/>
- YSO - Yleinen suomalainen ontologia; <http://finto.fi/yso/fi/>

8.1.2 Tieteenalakohtaiset, laajahko ala

- AFO - Luonnonvara- ja ympäristöontologia; <https://finto.fi/afo/fi/>
- Eduskunnan kirjaston asiasanasto – eduskuntatieto, oikeudellinen tieto, yhteiskuntatieto; <http://www.eduskunta.fi/kirjasto/EKS/>
- FinMeSH – Lääketieteen asiasanasto; <http://finto.fi/mesh/fi/>
- Helecon-asiasanasto – Kauppa- ja talousala; <http://web.lib.aalto.fi/fi/helevoc/>
- Kassu – Kasvien suomenkieliset nimet; <http://finto.fi/kassu/fi/>
- KULO – Kulttuurien tutkimuksen ontologia; <http://finto.fi/kulo/fi/>
- Paikkatieto-ontologia; <http://finto.fi/pto/fi/>
- TEPA-termipankki; <http://www.tsk.fi/tepa>
- TERO - Terveiden ja hyvinvoinnin ontologia; <http://finto.fi/tero/fi/>
- VALO – Valokuvausalan ontologia; <http://finto.fi/valo/fi/>

8.1.3 Tieteenalakohtaiset, suppeahko ala

- Hoidokki – Hoitotyön asiasanasto; <http://www.hoidokki.fi/>
- JUPO – Julkisen hallinnon palveluontologia; <http://finto.fi/jupo/fi/>
- KEKO - Kestävän kehityksen kasvatuksen ontologia; <http://finto.fi/keko/fi/>
- KULO - Kielitieteen ontologia; <http://finto.fi/kulo/fi/>
- Kitu – Kirjallisuudentutkimuksen asiasanasto; <http://www.finlit.fi/kitu>
- Lapponica-asiasanasto; <http://www.lapponica.net/sanasto1.htm>
- LIITO - Liiketoimintaontologia; <http://finto.fi/liito/fi/>
- Liikunnan ja urheilun asiasanasto; vain painettuna.
- MERO - Merenkulkualan ontologia; <http://finto.fi/mero/fi/>
- Pankki- ja rahoitussanasto; <http://www.tsk.fi/tsk/pankkisanasto/fi>
- Polas – Poliisin asiasanasto; <http://polas.eduix.fi:8080/polas/frameset.html>
- PUHO - Puolustushallinnon ontologia; <http://finto.fi/puhu/fi/>
- Teologian asiasanasto; <http://blogs.helsinki.fi/teolasiasanasto/>
- Tietotekniikan termitalkoot; <http://www.tsk.fi/tsk/termitalkoot/>



8.2 Ulkomaisia

- AAT – Art and Architecture Thesaurus; <http://www.getty.edu/research/tools/vocabularies/aat/index.html>
- ERIC – The Educational Resources Information Center; <http://eric.ed.gov/>
- LCSH – Library of Congress Subject Headings <http://authorities.loc.gov/>
- MeSH – Medical Subject Headings <http://www.nlm.nih.gov/mesh/>
- Svenska ämnesord <http://www.kb.se/katalogisering/Svenska-amnesord/>

8.3 Eurooppalaisia

- 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://ukdataservice.ac.uk/about-us/projects/cessda-elsst/details.aspx>



9 STANDARDIEN KUVAUKSET

9.1 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 ovat tietyin osin sovellettavissa myös muilla sektoreilla. Kirjastosektorilla kuvailusääntöjä on tulkittu hyvin yhteismitallisesti, mutta arkisto- ja museoaineistojen kuvailussa saattaa olla organisaatiokohtaisia eroja.

Kuvailevan metatiedon pohjana olevat formaatit ovat sektorikohtaisia ja osin entiteettikohtaisia. Kirjastot soveltavat MARC-formaattia, arkistot ISAAR CPF/EAC-CPF:ää; ISAD (G)/EAD:tä ja ISDF:ää, ja museot useita eri formaatteja, joista metatieto muunnetaan siirtoa varten LIDOon.

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 tiedonhakua. 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.

9.2 Rajapintastandardit

9.2.1 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 formateissa 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 haravoimiselle⁵⁵.

9.2.2 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 yhteydetön 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 määritellä 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 hakua tai estää sen kokonaan. Silloinkin, kun on olemassa standardirajapinta, tiedonhaku voi olla vaikeaa kohdetietokantojen semanttisten erojen vuoksi: tarvittava hakuterminä ei ole lainkaan, tai sitä ei ole indeksoitu.

Z39.50:ttä ja SRU:ta on pidetty kirjastojen standardeina. OpenSearch⁵⁶ on Amazonin kehittämä Web-pohjainen avoin hakurajapinta, josta yritettiin tehdä IETF-standardi⁵⁷, mutta hanke näyttää hyytyneen. Kuten edeltäjänsä, myöskään OpenSearch ei näytä pystyneen mainittavaan läpimurtoon, kuten ei myöskään Googlen Search Appliance⁵⁸. KDK:lle se ei sovellu, koska lähtökohtana ei ole metatiedon vaan kokotekstin indeksointi.

9.2.3 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 rajapintamäärittelyksiä, kuten ANSI/NISO:n Circulation Interchange Protocol, joka nimensä mukaisesti keskittyy lainaustietoihin.

9.3 Tunnukset

Asiakasliittymän ja PAS-palvelun tietosisältöjen haku- ja päivitysoperaatioiden on kohdistuttava oikein. Tämän varmistamiseksi jokaisella asiakasliittymässä ja PAS-palvelussa olevalla tietosisällöllä on oltava pysyvä, ainutkertainen ja toiminnallinen tunnus, jota voidaan käyttää tiedonhaussa ja josta voidaan muodostaa

⁵⁵ <http://www.niso.org/workrooms/resourcesync/>

⁵⁶ <http://www.opensearch.org>

⁵⁷ <https://tools.ietf.org/html/draft-ellermann-opensearch-03>

⁵⁸ <http://www.google.com/work/search/products/gsa.html>



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.

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

Suomessa käytössä olevilla standarditunnuksilla on yleensä tukiorganisaatio, kuten kansallinen ISBN-keskus⁵⁹, jolta voi pyytää apua tunnuksen soveltamista koskevissa ongelmatilanteissa.

9.4 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-palvelun 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. OAIS ei kuitenkaan pyri määrittelemään täsmällisesti tämän paketin rakennetta. Tästä vastaavat säiliöstandardit, joka KDK:ssa on METS. Tämän standardin avulla voidaan paketoita yhteen sekä itse dokumentti kaikkine tiedostoineen, että dokumenttiin liittyvät hallinnolliset ja kuvailevat metatiedot. Säiliöstandardit vastaavat tavaraliikenteen kontteja siinä, että ne helpottavat digitaalisten aineistojen käsittelyä.

METS-standardin valintaa tulee täydentää ohjeistuksella sen soveltamisesta. Kansallinen digitaalinen kirjasto -hanke ylläpitää ohjeistusta aineistojen paketoinnista⁶⁰.

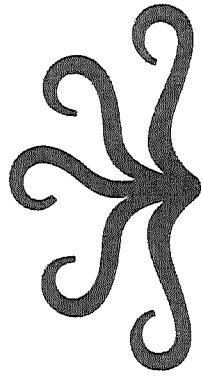
Pitkäaikaissäilytys edellyttää korkeatasoisia hallinnollisia metatietoja. Standardisalkku listaa käytettävät metatietoformaattit (kuten still-kuvien teknisen metatiedon MIXin), mutta näitä metatietoja sekä dokumenttien tiedostomuotoja ei esitellä standardisalkussa, vaan sitä tukevassa muussa ohjeistuksessa.

⁵⁹ <http://www.kansalliskirjasto.fi/julkaisuala/isbn.html>

⁶⁰ <http://www.kdk.fi/fi/pitkaaikaissailytys/maeairittely-ja-dokumentit/5-suomi/pitkaaikaissailytys/145-kdkn-hallinnolliset-ja-rakenteelliset-metatiedot-ja-aineiston-paketointi>



PAS-palvelun ja taustajärjestelmien väliseen kommunikointiin ei ole olemassa kansainvälistä standardia. ISO TC 46/SC 4 on aloittanut keväällä 2015 ranskalaiseen MEDONA-standardiin (Modélisation des échanges de données pour l'archivage) perustuvan standardointiprosessin, joka kestää pari vuotta. KDK:ssa on päätetty soveltaa tiedostojen siirrossa PAS-palveluun SFTP-standardia, joka on FTP:tä turvallisempi vaihtoehto. Muussa kommunikoinnissa PAS-palvelun ja taustajärjestelmien välillä käytetään HTTPS- ja REST-pohjaisia ratkaisuja, jotka minimoivat taustajärjestelmiin tehtävien muutosten tarpeen.



Kansallinen
digitaalinen
kirjasto