

DIGIME-STANDARDISALKKU

v 1.0

2021-02-05

Tämä määritys perustuu Opetus- ja kulttuuriministeriön
Kansallinen digitaalinen kirjasto -hankkeen standardisalkkuun



Lisenssi

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

Standardisalkun käyttäjälle annetaan oikeus jakaa eli kopiaida, levittää, näyttää ja esittää teosta sekä valmistaa muutettuja teoksia seuraavilla ehdoilla:

- *Digime-hanke 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	7
3.1	Kuvailusäännöt	7
3.2	Kuvailevan metatiedon formaatit.....	8
3.3	Henkilöiden, sukujen ja yhteisöjen tiedot	9
3.4	Sisällönkuvailun välineet	9
3.5	Hallinnollinen metatieto.....	10
3.6	Rakenteellinen metatieto	11
4	FINNAN JA TAUSTAJÄRJESTELMIEN RAJAPINNAT	12
4.1	Metatiedon haravointi Finnaan	12
4.2	Haku.....	12
4.3	Sovellustason standardoidut rajapinnat	13
5	TUNNISTEJÄRJESTELMÄT JA TUNNUKSET	14
5.1	Resoluutiopalvelut	14
5.2	Tunnisteet: aineistot ja niiden metatiedot.....	14
5.3	Tunnisteet: tekijät.....	15
5.4	Tunnisteet: paikat	15
6	AUTENTIKOINTI JA AUKTORISOINTI	16
6.1	Tiivistealgoritmit	16
7	PITKÄAIKAISSÄILYTYS JA AINEISTOJEN SIIRTO PAS-PALVELUUN	17
7.1	PAS-palvelun rajapinnat	17
7.2	PAS-palvelun auditointi	17
8	TÄRKEIMMÄT SANASTOT JA ONTOLOGIAT	18
8.1	Yleisluontoiset	18
8.2	Tieteenalakohtaiset, laajahko ala	18
8.3	Tieteenalakohtaiset, suppeahko ala.....	18
8.4	Ulkomaisia sanastoja	19

Kirjastojen, arkistojen ja museoiden yhteistyöstä ja yhteisien palvelujen toteutusta tukeva standardisalkku laadittiin alun perin osana vuonna 2017 päättyneen Kansallisen digitaalisen kirjaston eli KDK-hankkeen kokonaisarkkitehtuuria, joka on luettavissa osoitteessa <https://www.digime.fi/yhteentoimivuus/kokonaisarkkitehtuuri/>.

KDK-hankkeen päätyttyä standardisalkun ylläpidosta vastaa Digitaalisen kulttuuriperinnön kokonaisuuden (Digime, <https://www.digime.fi/>) tietoarkkitehtuuriryhmä.

KAM-sektorin standardisalkun ohella muistiorganisaatioiden on suotavaa soveltaa myös Julkishallinnon standardisalkun määrittämiä (katso <https://www.avoindata.fi/data/fi/dataset?q=standardisalkku> sekä <http://www.jhs-suositukset.fi/suomi/jhs181>).

1 JOHDANTO

Tässä standardisalkussa määritellään KAM-sektorin yhteisessä FINNA-hakuliitymässä ja pitkäaikaissäilytyksen Kulttuuriaineisto-PAS -palvelussa sekä muissa yhteisissä järjestelmissä kuten FINTO-ontologiapalvelussa sovellettavia standardeja. PAS-palvelun ja FINNAn osalta salkku täydentää niiden omia määrittämiä, jotka ovat käytettävissä osoitteissa <http://www.digitalpreservation.fi/specifications> ja <https://www.kiwi.fi/display/Finna/Ohjeet>.

Linjaukset koskevat myös kirjastojen, arkistojen ja museoiden omia tuotantojärjestelmiä siltä osin kuin niiden on oltava yhteismitallisia yhteisten järjestelmien kanssa. Yhteisen standardisalkun lisäksi sektoreilla voi olla omia standardisalkkuja ja muita ohjaavia dokumentteja kuten tietoarkkitehtuureja.

Digime-salkku sisältää jonkin verran perusteluja valinnoille, mutta ei anna yksityiskohtaisia palvelukohtaisia ohjeita standardien implementoinnista. Joidenkin standardien osalta on ollut tarpeen ja mahdollista viitata olemassa oleviin kansainvälisiin sovellusohjeisiin (profiileihin), ja muutamissa tapauksissa (esim. käyttöoikeuksien metadatan tallennus) on laadittu palvelukohtaista ohjeistusta (FINNA-ohjeistus osoitteessa <https://www.kiwi.fi/pages/viewpage.action?pageId=52298893>).

Standardisoinnin tavoitteena on varmistaa KAM-sektorin yhteisen järjestelmäkokonaisuuden toimivuus, joka edellyttää esimerkiksi eri organisaatioiden tallentaman metatiedon riittävää semanttista yhteismitallisuutta sekä yhteisiä ohjelmistorajapintoja. 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 edelleenkään tue edes kaikkia 1-tason standardeja; tästä aiheutuvat ongelmat tietojen siirrossa ja säilytyksessä ratkaistaan erikseen. Hankittavien uusien järjestelmien on tuettava ainakin välttämättömiä standardeja.

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 käytettävissä¹. Osa salkun standardeista on maksuttomia, osa maksullisia.

¹ Standardien avoimuudesta katso https://en.wikipedia.org/wiki/Open_standard

2 YLEISET STANDARDIT

FINNA-asiakasliittymän ja PAS-palvelun on tuettava UNICODE-merkkivalikoimastandardia² UTF-8 -koodauksella **(1)**. Standardi uudistetaan vuosittain, ja siihen lisättyjen uusien merkkien ja merkistöjen tukeminen on ajankohtaista sitten kun niiden tuki on lisätty käyttöjärjestelmiin ja varusohjelmistoihin.

Metatieto on suositeltavaa toimittaa taustajärjestelmistä Finnaan ja muihin sovelluksiin UTF-8 –koodattuina. **(2)**. Yksittäisten tietoelementtien merkkivalikoimasta voi olla olemassa tarkempia ohjeita. Esimerkiksi tunnuksissa on suositeltavaa käyttää vain tulostettavia ISO Latin 1 –merkkejä, vaikka tunnistejärjestelmän määrittäminen sallisi UNICODE-merkit.

PAS-järjestelmään toimitettavissa tekstidokumenteissa voi soveltaa UTF-8:n ohella myös ISO-8859-15, UTF-8, UTF-16 ja UTF-32 –koodausta **(3)**.

FINNA-asiakasliittymän on tuettava Web Content Accessibility Guidelines (WCAG) -suosituksen versiota 2.0³ tai versiota 2.1⁴ vähintään tason A vaatimusten mukaisesti⁵ **(1)**, suositeltavan tason ollessa AA **(2)**.

FINNAn käyttöliittymän suunnittelussa ja toteutuksessa sekä asiakkaille suunnatuissa ohjeissa on otettava huomioon EU:n saavutettavuusdirektiivin vaatimukset (ks. <https://www.saavutettavuusvaatimukset.fi/>).

² <http://unicode.org/>

³ <http://www.w3.org/TR/WCAG20>

⁴ <https://www.w3.org/TR/WCAG21/>

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

3 METATIEDOT

FINNA- ja PAS-palvelu edellyttävät riittäviä ja sisällöllisesti yhteneviä kuvailevia ja hallinnollisia metatietoja, joten metatietoformaateista ja kuvailun periaatteista sopiminen on välttämätöntä.

KAM-organisaatioiden yhteisten kuvailun linjausten tekeminen on mahdollista, koska sektoreiden käyttämät kuvailusäännöt ovat riittävän yhteismitallisia.

Metatieto voidaan jakaa perinteiseen kuvailevaan sekä hallinnolliseen ja rakenteiseen metatietoon. Hallinnollinen metatieto jakautuu edelleen tekniseen, käyttöoikeuksien ja pitkäaikaissäilytyksen metatietoon. Erityyppisille metatiedoille on omia formaatteja, joiden tietoelementit ovat osittain päällekkäisiä, esim. kirjastojen MARC 21 sisältää runsaasti myös hallinnollisen metadatan elementtejä⁶. Tekninen ja pitkäaikaissäilytyksen metadata tuotetaan pääosin ohjelmallisesti.

3.1 Kuvailusäännöt

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

Kuvailun kehittämisen perustaksi suositellaan kirjastojen RDA-säännöstöä (Resource Description and Access⁷) (2), vaikka sektoreilla tai yksittäisillä organisaatioilla voi lisäksi olla käytössä myös omia kuvailusäännöstöjä.

RDA tukee elektronisten aineistojen kuvailua sekä teosten ja niiden manifestaatioiden linkittämistä toisiinsa. RDA on julkaistu 2011, ja ensimmäiset kirjastot ottivat sen käyttöön jo vuonna 2013. Semanttisen yhteismitallisuuden parantamiseksi arkistojen ja museoiden toimijakuvailun kehittämisessä otetaan huomioon RDA:n toimijakuvailuohjeistus.

Suomalainen versio RDA:sta julkaistiin 2015, ja sitä alettiin soveltaa kirjastosektorilla vuoden 2016 alussa. Käännöstyöstä ja RDA:n käyttöönotosta Suomessa on vastannut Kansalliskirjasto. Kansallisen RDA-sovelluksen teossa on pyritty ottamaan Finnan ja eri KAM-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).

Erityisarkistoilla on omat kansainvälisesti yhdenmukaiset käytäntönsä (esim. data-arkistot ja elokuva-arkistot).

ICA:n EGAD-työryhmä valmistelee nykyiset standardit korvaavaa arkistokuvailun kansainvälistä käsitelmää RiC-CM "Records in Contexts" ja sitä tarkentavaa ontologiaa RiC-O.

Kansallisella tasolla on julkaistu luonnos em. voimassaoleviin arkistokuvailustandardeihin sekä RDA:han ja ISO 21127 *A Reference ontology for the interchange of cultural heritage information* (2006, uusin versio 2014) standardiin perustuva "Arkistokuvailun kansallinen käsitelmä". Uusien yleisten kansallisten arkistokuvailusääntöjen laadintaan ryhdytään sen jälkeen, kun ICA:n tämän hetkiset kuvailusäännöt korvautuvat RiC-CM mallilla. Nämä säännöt tulevat korvaamaan vuonna 1997 julkaistut *Arkistojen kuvailu-*

⁶ <http://www.loc.gov/marc/bibliographic/mapping/index.html>

⁷ <http://rdatoolkit.org/>

ja luettelointisäännöt. Ahaa -hankkeen kuvailuohjetta voidaan hyödyntää kansallisten sääntöjen tuottamisvaiheessa.

Syntysähköisen asiakirja-aineiston muodostumisen yhteydessä tallennettava metatieto noudattaa Kansallisarkiston Sähke2-metatietomallia. Sen versio 1.3 ilmestyi maaliskuussa 2020⁸. Tiedonhallintalain voimaantulon myötä arkistoaineiston säilytysvaihe ja arkistovaihe erottuvat mallissa aiempaa selkeämmin toisistaan. FINNAan siirrettäessä Sähke2-mallin mukainen metatieto muunnetaan EAD-formaattiin. PAS-palveluun siirretään sekä EAD3- että Sähke2-metatieto.

Museosektorilla on vuodesta 2014 alkaen sovellettu Iso-Britanniassa Collections Trust -säätiön laatimaa ja ylläpitämää SPECTRUM-luettelointistandardia, jonka viimeisin versio on vuonna 2017 ilmestynyt 5.0⁹. SPECTRUMin pohjalta on laadittu ohjeet esineiden, arkistoaineiston, av-aineistojen, taideteosten ja valokuvien kuvailuun¹⁰.

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. FINNA- ja PAS-järjestelmissä sekä Formula-palvelussa metatiedot tallennetaan ilman konversioita siinä muodossa, jossa ne on vastaanotettu. Kunkin formaatin tietoelementtien indeksointiratkaisut on kehitettävä FINNAssa erikseen. Tämä onnistuu helpoimmin ja palvelu toimii tehokkaimmin 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 FINNAan tulisi soveltaa jotakin seuraavista formaateista (2):

- MARC 21 (MACHINE-Readable Cataloguing)
- Dublin Core (15 kentän perus-DC, Qualified Dublin Core sekä julkaisuarkistojen tekstiaineistojen metadatasuositus)
- MODS (Metadata Object Description Schema), versio 3.7 tai aiempi
- EAD3 (Encoded Archival Description), versio 1.1.1 tai aiempi
- EAD 2002 (vanha versio EAD3a suositellaan)
- EAC-CPF (Encoded Archival Context- Corporate bodies, Persons and Families)
- EBUCore (PAS-palvelussa)
- LIDO (Lightweight Information Describing Objects), versio 1.0 (uusi versio tulossa)
- VRA Core (Visual Resources Association Core Categories), versio 4.0
- FORWARD data schema (based on CEN 15907 Film identification - Enhancing interoperability of metadata - Element sets and structures)¹¹
- DDI (Data Documentation Initiative), versio DDI Lifecycle 3.3 tai aiempi, DDI Codebook 2.5.1 tai aiempi
- DataCite Metadata Schema, versio 4.3 tai aiempi

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

⁸<https://arkisto.fi/fi/viranomaisille/Julkishallinnon-asiakirjahallinnon-ja-arkistotoimen-ohjaus/maeraeykset/saehke2-maeraeays>

⁹ <https://collectionstrust.org.uk/spectrum/>

¹⁰ <https://www.museovirasto.fi/fi/palvelut-ja-ohjeet/museoalan-kehittaminen>

¹¹ <https://www.kiwi.fi/download/Agent+Work+Information+for+FORWARD>

dokumentaation avulla. Käytännön toteutuksesta sovitaan tarpeen mukaan erikseen. Siirtoformaattien tarkemmat sovellusohjeet sovitaan niitä hyödyntävillä sektoreilla ja formaattien indeksointiperiaatteista sovitaan FINNA-palvelun kanssa erikseen.

Dublin Coren osalta sallitaan aineistokohtaisten sovellusprofiilien kuten julkaisuarkistojen tekstiaineistojen metadatasuosituksen¹² käyttö, mutta niiden indeksointiperiaatteet on sovittava FINNA-palvelun kanssa erikseen.

DataCite-määrittystä sovelletaan tutkimusaineistojen kuvailussa.

Lisätietoja

FINNA-palvelu on luonut tiedontuottajien kanssa taulukon kuvailevan metatiedon formaattien kenttien mappauksesta FINNAn hakuindeksiin¹³. Siitä puuttuu muutamia käytössä olevia formaatteja.

FINTO-palvelussa on keskusteltu mahdollisuudesta luoda eri metadataformaattien kenttien suhteita kuvaavasta Vocabulary Mapping Framework –ontologiasta¹⁴ suomalainen versio. Se tukisi FINNA-palvelun mappauksia ylläpitäen ja mahdollista laajentamista uusiin formaatteihin.

3.3 Henkilöiden, sukujen ja yhteisöjen tiedot

Henkilöt, suvut ja yhteisöt kuvaillaan kirjasto-, arkisto- ja museosektoreilla yhteisten tietoelementtien osalta RDA-sääntöjen mukaisesti. Metatiedot suositellaan tallennettavaksi niin kattavina, että ne mahdollistavat yksiselitteisen tunnistamisen **(2)**. Mahdollisen yhteisen nimitietopalvelun sisäinen 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)
- FORWARD data schema (based on CEN 15907 Film identification - Enhancing interoperability of metadata - Element sets and structures)

Konversiot näistä formaateista tulevassa 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 -määrittämiä¹⁵.

Kansainvälisen ISNI-tietokannan rajapinnoista päätetään erikseen.

3.4 Sisällönkuvailun välineet

Sisällönkuvailussa tulee käyttää FINTO-palvelun¹⁶ Yleistä suomalaista ontologiaa tai siihen perustuvia erityisalojen ontologioita **(1)**. Niiden rinnalla/täydennykseksi voidaan aihealueen niin vaatiessa soveltaa kansainvälisiä sanastoja tai luokituksia. Esimerkkejä näistä järjestelmistä ovat UDK, Yleisten kirjastojen kymmenluokitus (YKL), Outline of Cultural Materials (OCM), Iconclass ja Classification of Individual Consumption according to Purpose (COICOP).

¹² <https://www.kiwi.fi/display/Julkaisuarkistopalvelut/Metadatasuositus+julkaisuarkistojen+tekstiaineistolle>

¹³ <https://www.kiwi.fi/display/Finna/Kenttien+mappaukset+eri+formaateista+Finna+indeksiin>

¹⁴ <https://www.doi.org/VMF/index.html>

¹⁵ https://joinup.ec.europa.eu/asset/core_vocabularies/description

¹⁶ <http://finto.fi/fi/>

Sanastojen kuvaamiseen käytetään FINTOssa SKOS-XL -määrittystä (Simple Knowledge Organization System eXtension for Labels¹⁷), jota laajennetaan tarpeen mukaan. Ontologioiden kuvaamisessa sovelletaan lisäksi RDFS (RDF Schema) ja OWL2 (Web Ontology Language, versio 2) -ontologiakielimäärittäjiä.

3.5 Hallinnollinen metatieto

Kuvailusäännöt ja kuvailevan metatiedon formaatit eivät kata kaikkea pitkäaikaissäilytyksessä tarvittavaa metatietoa. CSC:n PAS-palvelua varten on sen vuoksi laadittu ohje pitkäaikaissäilytyksessä tarvittavista hallinnollisista metatiedoista ja niiden tallentamisesta PAS-palveluun tallennettaviin siirtopaketteihin¹⁸. Kaikki PAS-palveluun siirrettävä aineisto on varustettava ohjeen mukaisin hallinnollisin metatiedoin **(1)**.

Koska kattavia hallinnollisen metadaan kuvailusääntöjä ei ole, PAS-ohjeistuksella on merkittävä rooli kuvailun minimitason määrittelyssä. KAM-organisaatiot voivat tallentaa siirtopaketteihin myös muuta, esimerkiksi digitointiprosessissa syntyneitä metatietoa.

Teknisen metatiedon de facto -standardiformaatteja on käytettävissä yksi kutakin aineistotyyppiä kohden.

Tekninen metatieto

Vaaditaan seuraavien metatietoformaattien käyttöä soveltuvin osin **(1)**:

- MIX (NISO Metadata for images in XML Schema)¹⁹
- ADDML (Archival Data Description Markup Language)²⁰
- AudioMD (Technical metadata for Audio)²¹
- VideoMD (Technical metadata for Video)

Pitkäaikaissäilytyksen metatieto

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

- PREMIS Data dictionary for preservation metadata, versio 2.2.

PREMIS-formaatilla kuvattavat pitkäaikaissäilytykseen liittyvät tapahtumat määritellään PAS-palvelua ja KAM-organisaatioiden tuotantojärjestelmiä varten erikseen. Tapahtuma voi olla esimerkiksi tiedoston kopiointi toiselle tallenteelle tai tiedostoformaatin migraatio.

Käyttöoikeustiedot

PAS-palvelu

PAS-palvelua koskevat käyttöoikeustiedot – kuten sen kuvaaminen, keillä on oikeus tehdä migraatio arkistoidulle dokumentille - tallennetaan PREMIS-metatietoformaatin²² Rights-osioon erikseen sovittavalla tavalla **(1)**. Tapahtumakuvauksista tulee ilmetä toimija (henkilö / ohjelmisto), ajankohta sekä tulos.

Tuotantojärjestelmät

Tuotantojärjestelmiin 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)**.

Käyttöoikeuksien metatietojen (pääsyn ja käytön ehdot, oikeudet ja rajoitukset) tallennuksessa tulee käyttää luvussa 3.2 listattuja formaatteja **(1)**.

¹⁷ <https://www.w3.org/TR/skos-reference/skos-xl.html>

¹⁸ <http://www.digitalpreservation.fi/specifications>

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

²⁰ <http://www.arkivverket.no/arkivverket/Arkivbevaring/Elektronisk-arkivmateriale/Standarder/ADDML>

²¹ <http://www.loc.gov/standards/amdvmd/>

²² <http://www.loc.gov/standards/premis/>

Kirjastoille on laadittu ohje käyttöoikeuksien metatietojen tallentamiseen MARC 21 -formaatussa²³. Sitä tulisi soveltaa ohjetta Finnaan menevän käyttöoikeuksien metatiedon tallennuksessa **(2)**.

Lisenssin ja muiden sopimusten sisältö voi olla vaikeasti kuvattavissa metatiedoissa; suositeltava ratkaisu on tallentaa metatietoihin vain lisenssin nimi ja linkki verkossa olevaan lisenssiin tai sen tiivistelmään. Jotta tämä olisi mahdollista, tulisi aineistoille käyttää Creative Commons lisenssejä, Rights statementsejä²⁴, tai muita yleisesti tunnettuja lisenssejä **(2)**.

FINNA-palvelussa käytettävät tiedot kuten lisenssin nimi tulisi tallentaa yhteismitallisesti riippumatta siitä, mitä formaattia käytetään **(2)**.

Käyttörajoitusten (aineistot ja niiden metatiedot) kuvaukseen tulisi sisältyä peruste (laki tai sopimus), lain / sopimuksen nimi, käyttörajoituksen selite sekä rajoituksen kesto **(2)**.

Tiedostomuodot

PAS-palveluun tallennettavissa tai siirrettävissä olevat säilytys- ja siirtokelpoiset tiedostomuodot määritellään vuosittain tarkistettavassa ohjeessa²⁵.

PAS-palveluun lähetettävään siirtopakettiin voidaan tallentaa myös alkuperäinen tiedostomuoto autenttisuuden takaamiseksi, vaikka se ei olisi siirto- eikä säilytyskelpoinen. Nämä tiedostot on koodattava siirtopaketissa siten, ettei niitä validoida PAS-palvelussa säilytyspakettia muodostettaessa. PAS-palvelu säilyttää nämä tiedostot vain bittitasolla.

3.6 Rakenteellinen metatieto

PAS-järjestelmään siirrettävien dokumenttien rakenteen kuvaamiseen on käytettävä joko METS-standardin StructMap -osiota tai tiedostomuodon sisäistä rakenteisen metadatan tallennusta **(1)**.

²³ <https://www.kiwi.fi/pages/viewpage.action?pageId=37979531>

²⁴ <https://rightsstatements.org/fi/>

²⁵ <http://www.digitalpreservation.fi/specifications>

4 FINNAN JA TAUSTAJÄRJESTELMIEN RAJAPINNAT

4.1 Metatiedon haravointi Finnaan

Kaikista tietojärjestelmistä pitää olla mahdollista haravoida kuvailevat metatiedot FINNAan.

Metatietojen haravointiin sovelletaan OAI-PMH -protokollan (The Open Archives Initiative Protocol for Metadata Harvesting) versiota 2.0²⁶ **(1)**. OAI-PMH:n soveltamisesta on laadittu erillinen ohje²⁷.

Haravoinnissa käytetään RecordManager-sovellusta, joka tukee seuraavia metatietoformaatteja:

- Dublin Core
- Qualified Dublin Core
- EAC-CPF
- EAD3
- EAD 2002 (vanhentunut, suositellaan EAD3)
- EDM
- FORWARD
- LIDO
- MARCXML (MARC21)

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

4.2 Haku

Finna tarjoaa ohjelmistorajapinnan, joka mahdollistaa helpon tavan tehdä hakuja ja poimia metatietoja Finnasta. Finna API:n englanninkielinen kuvaus löytyy osoitteesta <https://www.kiwi.fi/pages/viewpage.action?pagelId=53839221>. (Katso myös api.finna.fi.)

Melinda-yhteisluettelosta pitää voida tehdä Z39.50- ja SRU-hakuja **(1)**. Näiden rajapintojen kuvaus löytyy osoitteesta <https://www.kiwi.fi/pages/viewpage.action?pagelId=166101210>

DSpace-julkaisuarkistosovellus tukee OpenSearch-standardia. Doria-julkaisuarkiston OpenSearch-rajapinnan dokumentaatio löytyy osoitteesta <https://www.kiwi.fi/pages/viewpage.action?pagelId=45782169>

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)**.

PAS-järjestelmän hakutoiminnot on dokumentoitu rajapinnan kuvauksessa³¹.

²⁶ <http://www.openarchives.org/OAI/openarchivesprotocol.html>

²⁷ <https://www.kiwi.fi/display/Finna/Finna+ja+OAI-PMH>

²⁸ <http://www.niso.org/standards-committees/ncip>

²⁹ <http://www.w3.org/TR/soap/>

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

³¹ <http://www.digitalpreservation.fi/specifications/interface-history>

4.3 Sovellustason standardoidut rajapinnat

Integroitujen kirjastojärjestelmien pitää **(1)** tarjota FINNAlle Digital Library Foundationin ILS-DI -standardin mukainen tiedonsiirto- ja palvelurajapinta³². Muut taustajärjestelmät voivat **(3)** tukea ILS-DI:tä.

³² http://old.diglib.org/architectures/ilsdi/DLF_ILS_Discovery_1.0.pdf

5 TUNNISTEJÄRJESTELMÄT JA TUNNUKSET

PAS-palveluun tallennettaville pitkäaikaissäilytettävälle elektronisille aineistoille on annettava manifestaatiotason standarditunnus **(1)**. Tämän lisäksi aineistolla voi olla teostason tunnus **(2)**.

Näiden tunnusten pitää olla toiminnallisia eli PID-tunnuksia (persistent identifier) tai ne pitää voida muuntaa sellaisiksi **(1)**. Hyvin hallintoitujen PID-tunnistejärjestelmien kuten DOI:n ja URN:n käyttöä suositellaan **(2)**.

PID-tunnukset tulee esittää hyperlinkkeinä (HTTP URI) **(1)**.

Jos elektronisen aineiston kuvailutiedot tulevat haettaviksi FINNA-palvelussa mutta aineisto ei ole pitkäaikaissäilytettävää, aineistolle on annettava standarditunnus **(1)**, jonka esittämistä PID-tunnuksena suositellaan **(2)**.

Aineiston kaikilla kappaleilla tulisi olla sama sijaintipaikasta riippumaton PID-tunnus. Resolverin tulisi linkittää PID-tunnus aineiston kappaleiden sijaintipaikkoihin eli URL-osoitteisiin siten, että asiakas näkee vain ne kappaleet, jotka ovat hänen käytettävissään **(2)**.

5.1 Resoluutiopalvelut

PID-tunnuksien avulla pitää tarjota resoluutiopalveluja, eli esim. linkitys identifioidun aineiston yhteen tai useampaan kopioon tai esim. aineiston metatietoihin **(1)**.

Resoluutiopalvelun tarjoaja ja käytettävissä olevat palvelut riippuvat PID-järjestelmästä.

DOI-palveluja tarjoavat Suomessa tätä kirjoitettaessa (2020-12) CSC – Tieteen tietotekniikan keskus Oy tutkimusaineistoille ja Tieteellisten seurojen valtuuskunta TSV Journal.fi –palvelussa julkaistaville artikkeleille.

URN-resoluutiopalvelua tarjoaa Kansalliskirjasto.

Resoluutiopalvelimien tarjoama toiminnallisuus riippuu pitkällä aikavälillä e-aineistojen hallinnointiin käytettävistä sovelluksista (esim. PAS-palvelu ja kirjastojen julkaisuarkistot), PID-tunnistejärjestelmästä sekä PID-resolverin tekniikasta. Toistaiseksi sekä DOI- että URN-järjestelmät tarjoavat vain linkityksen yhteen tai useampaan aineiston kopioon tai metatietoihin.

5.2 Tunnisteet: aineistot ja niiden metatiedot

Aineistoille tulee käyttää niille soveltuvia standarditunnuksia aina kun sopiva tunnistejärjestelmä on olemassa **(1)**. Esim. kirjalle tulee antaa ISBN-tunnus.

Jos PAS-palveluun tallennettavalle aineistolle ei ole olemassa aineistokohtaista standarditunnistetta, identifiointiin suositellaan käytettäväksi joko DOI-tunnistetta tai Kansalliskirjaston hallinnoimaa URN:NBN-tunnistetta (National Bibliography Number) **(2)**.

Handle-tunnuksia saa käyttää, jos URN- ja DOI-vaihtoehdot eivät tule kyseeseen **(3)**. Muita PID-vaihtoehtoja (ARK, PURL) tulee välttää.

URN:NBN-tunnuksia suositellaan aineistojen metatietojen identifiointiin **(2)**.

HUOM URL-osoitteita ei saa käyttää tunnuksina PAS-järjestelmään tallennettaville aineistoille eikä niiden metatiedoille, koska osoitteiden toimivuutta ja ainutkertaisuutta ei voida taata. Aineiston alkuperäisestä osoitteesta voi olla mahdotonta tehdä nimipalvelun avulla uudelleenohjaus PAS-palveluun, ja PAS-palveluun voidaan myöhemmin lähettää muuttunut tiedosto jonka URL-osoite on sama kuin PAS-palvelussa jo olevan tiedoston.

5.3 Tunnisteet: tekijät

Henkilö- ja yhteisötekijöiden tunnistamiseen on aktiivitutkijoita lukuun ottamatta suositeltava käyttää International Standard Name Identifier (ISNI) –standarditunnistetta (ISO 27729)³³ **(2)**.

Kansalliseksi tutkijatunnisteeksi suositellaan ORCID-tunnusta³⁴ **(2)**. Se vastaa rakenteeltaan ISNIä, mutta sillä on eri ylläpito-organisaatio. Vastuu ORCID-tunnusten luonnista ja niihin liittyvien metatietojen tallentamisesta ja ylläpidosta on hajautettu tutkijoille, kun ISNI-järjestelmässä vastuu on keskitetty kansainväliselle ISNI-keskukselle ja sen Registration Agency – eli RAG-yhteistyöorganisaatioille.

Tutkimusorganisaatioille voidaan käyttää ISNI:n ohella myös ROR-tunnusta **(3)**.

Kansalliskirjasto on saanut ISNI RAG –statuksen syksyllä 2020. Kansallisen ORCID-konsortion koordinaattori on CSC.

Yhteisten nimitietojärjestelmien kuten Finton Kanto-ontologian³⁵ (Kansallinen toimijatiedot) sekä sektorikohtaisten järjestelmien kuten kirjastojen Asterin on tuettava sekä ORCID- että ISNI-tunnuksia **(1)**.

5.4 Tunnisteet: paikat

Paikkatietojen identifiointiin tulee soveltaa INSPIRE-direktiivin mukaisia, Maanmittauslaitoksen ylläpitämiä yksilöiviä tunnuksia³⁶ **(1)**.

³³ <http://isni.org/>

³⁴ <https://tutkijatunniste.fi/>

³⁵ <https://finto.fi/finaf/fi/>

³⁶ <http://docs.jhs-suositukset.fi/jhs-suositukset/JHS193/JHS193.pdf>

6 AUTENTIKOINTI JA AUKTORISOINTI

Varmenteiden hankinnassa ja käsittelyssä sekä sähköisiin allekirjoituksiin liittyvien avainten hallinnassa on otettava huomioon laki vahvasta sähköisestä tunnistamisesta ja sähköisistä luottamuspalveluista 7.8.2009/617³⁷, sitä päivittävä laki 533/2016³⁸ ja näiden lakien 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-protokollan³⁹ mukaisia palvelinvarmenteita **(1)**.

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

Heikon identiteetin tunnistautumispalveluissa tulisi käyttää OpenID:tä **(2)**.

Tunnistautumispalveluilla tulisi oltava käytettävissä joko organisaatiokohtainen ratkaisu tai jokin julkisen sektorin tunnistautumisfederaatioiden palvelu, esimerkiksi Haka (Shibboleth/SAML) **(2)**, VIRTU (SAML 2.0) **(2)** tai VETUMA (SAML 2.0) **(2)**.

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 tiiviste mahdollisimman varhaisessa vaiheessa, mutta viimeistään ennen PAS-järjestelmään lähettämistä. Laskennassa tulee käyttää MD5⁴⁰-, SHA-1⁴¹-, SHA-224⁴², SHA-256⁴³, SHA-384⁴⁴ tai SHA-512⁴⁵ algoritmia **(1)**.

³⁷ <http://www.finlex.fi/fi/laki/ajantasa/2009/20090617>

³⁸ <https://finlex.fi/fi/laki/alkup/2016/20160533>

³⁹ <https://tools.ietf.org/html/rfc4252>

⁴⁰ <http://fi.wikipedia.org/wiki/MD5>

⁴¹ <http://fi.wikipedia.org/wiki/SHA-1>

⁴² <http://en.wikipedia.org/wiki/SHA-224>

⁴³ <http://en.wikipedia.org/wiki/SHA-256>

⁴⁴ <http://en.wikipedia.org/wiki/SHA-384>

⁴⁵ <http://en.wikipedia.org/wiki/SHA-512>

7 PITKÄAIKAISSÄILYTYS JA AINEISTOJEN SIIRTO PAS-PALVELUUN

Pitkäaikaissäilytyksen viitemallina sovelletaan CCSDS:n (The Consultative Committee for Space Data Systems) julkaisemaa OAIS-viitemallia⁴⁶ (Open Archival Information System), joka on julkaistu myös ISO-standardina ISO 14721:2012 (1). Standardista on suomenkielinen versio Viitemalli pitkäaikaissäilytysarkistolle (SFS 5972 2015).

OAIS-viitemallissa pitkäaikaissäilytykseen siirrettävää yksittäistä aineistokokonaisuutta kutsutaan siirtopaketiksi. Paketti sisältää sekä itse aineiston että sitä kuvaavan hallinnollisen ja kuvailevan metatiedon. ”Paketointiin” tulee soveltaa METS (Metadata Encoding and Transmission Standard) -standardin⁴⁷ versiota 1.11 (1).

7.1 PAS-palvelun rajapinnat

PAS-palvelussa säilytettävien aineistojen hakuun ja siirtoon sovelletaan CSC:n ylläpitämää rajapintaa, jonka dokumentaatio löytyy osoitteesta <http://urn.fi/URN:NBN:fi-fe2020100578097>. Käyttäjäorganisaatioiden tuotantojärjestelmien ei tarvitse tukea sitä.

PAS-hankkeessa on keskusteltu SWORD 3.0 –protokollaan⁴⁸ perustuvan standardirajapinnan toteuttamisesta. Tämä rajapinta olisi toiminnallisesti nykyistä kehittyneempi, mutta edellyttäisi tukea myös PAS-palvelun käyttäjäorganisaatioiden tuotantojärjestelmissä ja FINNAssa.

7.2 PAS-palvelun auditointi

PAS-palvelun tietoturva on auditoitava ISO 27001 –standardia soveltaen (1). Palvelu läpäisi auditoinnin vuonna 2019.

⁴⁶ <http://public.ccsds.org/publications/archive/650x0m2.pdf>

⁴⁷ <http://www.loc.gov/standards/mets/>

⁴⁸ <https://swordapp.github.io/swordv3/swordv3.html>

8 TÄRKEIMMÄT SANASTOT JA ONTOLOGIAT

Sisällönkuvailussa suositellaan käytettäväksi FINTO-palvelun sanastoja ja ontologioita. Kuvailussa voi lisäksi käyttää muita koti- ja ulkomaisia sanastoja.

Alla esimerkkejä KAM-sektorilla käytetyistä sanastoista katteen mukaan jaoteltuna.

8.1 Yleisluontoiset

- JUHO – Julkishallinnon ontologia; <http://finto.fi/juho/fi/>
- Kauno – Fiktiivisen aineiston ontologia; <http://finto.fi/kauno/fi/>
- KANTO – Kansalliset toimijatiedot; <https://finto.fi/finaf/fi/>
- KOKO-ontologia; <http://finto.fi/koko/fi/>
- MAO/TAO – Museoalan ja taideteollisuusalan ontologia; <http://finto.fi/maotao/fi/>
- Metatietosanasto; <https://finto.fi/mts/fi/>
- MUSO – Musiikin ontologia; <http://finto.fi/muso/fi/>
- SLM – Suomalainen lajityyppi- ja muotosanasto; <http://finto.fi/slm/fi/>
- YSO - Yleinen suomalainen ontologia; <http://finto.fi/ys0/fi/>
- YSO-paikat; <http://finto.fi/ys0-paikat/fi/>

8.2 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/>
- GEO – Geologian ontologia; <http://finto.fi/geo/fi/>
- KASSU – Kasvien suomenkieliset nimet; <http://finto.fi/kassu/fi/>
- KULO – Kulttuurien tutkimuksen ontologia; <http://finto.fi/kulo/fi/>
- PTO - Paikkatieto-ontologia; <http://finto.fi/pto/fi/>
- TERO - Terveiden ja hyvinvoinnin ontologia; <http://finto.fi/tero/fi/>
- VALO – Valokuvausalan ontologia; <http://finto.fi/valo/fi/>

8.3 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/>
- KTO - Kielitieteen ontologia; <http://finto.fi/kto/fi/>
- KITO – Kirjallisuudentutkimuksen ontologia; <http://finto.fi/kito/fi/>
- Lapponica-asiasanasto; <http://finto.fi/lapponica/fi/>
- LIIKO – Liikenteen ontologia; <http://finto.fi/liiko/fi/>
- LIITO - Liiketoimintaontologia; <http://finto.fi/liito/fi/>
- Liikunnan ja urheilun asiasanasto; vain painettuna.
- MERO - Merenkulkualan ontologia; <http://finto.fi/mero/fi/>
- OIKO – Oikeushallinnon ontologia; <http://finto.fi/oiko/fi/>
- PUHO - Puolustushallinnon ontologia; <http://finto.fi/puho/fi/>
- SEKO – Suomalainen esityskokoonpanosanasto; <http://finto.fi/seko/fi/>
- SOTO – Sotatieteen ontologia; <http://finto.fi/soto/fi/>

8.4 Ulkomaisia sanastoja

- 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 <https://metadatabyran.kb.se/amnesord-och-genre-form/svenska-amnesord>

- EUROVOC; <http://eurovoc.europa.eu/>
- 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 <https://elsst.ukdataservice.ac.uk/thesaurus-search.aspx>
- DDI Vocabularies; <https://ddialliance.org/controlled-vocabularies>
- CESSDA Topic Classification; <https://vocabularies.cessda.eu/vocabulary/TopicClassification?lang=en>