

Oikeuksien hallintaan liittyvät metatiedot -selvitys

Kansallinen
digitaalinen
kirjasto



AVOIN TIEDE
JA TUTKIMUS



Sisällys


1. Johdanto	5
Käsitteet.....	7
2. Oikeuksien hallinta ja metatiedot.....	8
2.1 Lisenssien kuvailu.....	11
2.2. Pitkäaikaissäilytys ja käyttöoikeustiedot.....	11
2.3 Rajapinnat eri järjestelmien välillä.....	12
3. Käyttöoikeuksien kuvaaminen eri metadastandardeissa	14
3.1 Kansainvälinen yleinen kehitys	14
DCAT.....	14
Europeana	15
RDA ja CODATA	16
Europeana Data Model	17
DataCite ja NISO.....	17
4. Suositukset oikeuksien hallintaan liittyvien metatietojen luomiseen ja käyttöön	19

Liite 1. Metatietoformaattit ja esimerkit eri järjestelmistä

Liite 2. Oikeuksiin liittyviä luokituksia metatiedoissa

Liite 3. Käsitteitä



Julkaisu Oikeuksien hallintaan liittyvät metatiedot -selvitys	
Julkaisija Avoin tiede ja tutkimus -hanke	Julkaisuajankohta xx.2.2017
Tekijä Avoin tiede ja tutkimus -hanke	
Lisenssi  Tämä teos on lisensoitu Creative Commons Nimeä 4.0 Kansainvälinen -lisenssillä .	
Julkaisun jakelu avointiede.fi ; http://urn.fi/URN:NBN:fi-fe201702101528	
Yhteystiedot http://avointiede.fi/ avointiede@postit.csc.fi	

Työryhmän työskentelyyn ja tämän selvityksen kirjoitusprosessiin ovat osallistuneet seuraavat henkilöt:

Ludvig Forsman, Turun yliopiston kirjasto
Miikka Haimila, Museovirasto
Juha Hakala, Kansalliskirjasto
Sampsa Heinonen, Museoliitto
Heikki Helin, CSC
Marja Hjelt, Yleisten kirjastojen keskuskirjasto
Juha Henriksson, Musiikkiarkisto JAPA
Pekka Henttonen, Tampereen yliopisto, Informaatiotiede
Outi Hupaniittu, Suomalaisen Kirjallisuuden Seura
Nina Hyvönen, Kansalliskirjasto
Jyrki Ilva, Kansalliskirjasto
Anu Jäppinen, Yleisten kirjastojen keskuskirjasto
Hanna Koivula, Luonnontieteellinen keskusmuseo LUOMUS
Kimmo Koskinen, Helsingin yliopiston kirjasto
Mikko Kuutti, Kavi
Tommy Lahtinen, Åbo Akademis bibliotek
Mietta Lennes, FIN-CLARIN, Helsingin Yliopisto
Ismo Malinen, Museovirasto
Markus Merenmies, Kansallisarkisto
Maria Miinalainen, Svenska litteratursällskapet i Finland
Katja Moilanen, Yhteiskuntatieteellinen tietoarkisto
Piia Naukkarinen, Kansalliskirjasto
Suvi Remes, CSC (käsitteet)
Seija Saari, Kansallisarkisto
Jukka Saarinen, Suomalaisen Kirjallisuuden Seura
Juha Törnroos, CSC
Maria Virtanen, Kansalliskirjasto
Marjut Vuorinen, THL
Stina Westman, CSC
Joonas Nikkanen, CSC
Jessica Parland-von Essen, CSC (puheenjohtaja)



Tietoa on lisäksi kerätty monilta eri tahoilta, jotka ovat ystävällisesti auttaneet ja antaneet tietoa.



1. Johdanto

Oikeuksien hallinnan metatiedot kertovat asiakkaille kuvailtavien aineistojen saatavuudesta ja käyttömahdollisuuksista. Niitä voidaan käyttää hakukriteereinä; asiakas voi esimerkiksi rajata haun pelkästään sellaiseen tutkimusaineistoon joka on vapaasti käytettävissä joko Public Domain -merkinnän tai CC0-lisenssin perusteella, tai julkaisuihin joilla ei ole pääsyräjoituksia. Tehokas hakukäyttö edellyttää kuitenkin metatietojen yhteismitallisuutta. Aineistojen saatavuutta voidaan myös hallita riittävien metatietojen avulla.

Oikeuksien hallinnan metatietoja ei ole ollut tarpeen tallentaa perinteiselle kirjasto- ja arkistoaineistoille, mutta digitaalisille tietoaineistoille ne ovat välttämättömiä, koska asiakkaiden täytyy tietää, mitä he voivat aineistoille tehdä kopioituaan ne omalla työasemalleen tai heidän on saatava tietää miksi kopiointi ei ole mahdollista. Oikeuksien hallinnan metatietojen tallennusta ei kuitenkaan tehdä vielä kattavasti eivätkä käytännöt ole täysin vakiintuneita.

Oikeuksien hallintaan liittyvät metatiedot -työryhmä on kevään 2016 aikana tehnyt suhteellisen laajan taustoittavan kartoituksen olemassa olevista arkkitehtuureista KDK:n sekä ATT:n piirissä. Lisäksi on käyty läpi keskeiset käyttöoikeuksien metadastandardit ja alan kansainvälinen kehitys. Työssä lähdettiin valmistelussa tunnistetuista ongelmista, kuten yhteisten käsitteiden ja selkeän ohjeistuksen puutteesta, jotka ovat johtaneet ongelmiin käytettävyyden ja yhteentoimivuuden osalta. Pahimmassa tapauksessa suuria määriä aineistoja jää käytön saavuttamattomiin koska sovellukset eivät pysty tulkitsemaan tallennettuja metatietoja tai ne puuttuvat kokonaan.

Samalla kartoitettiin tarpeita ja ongelmakohtia. Laajapohjainen keskustelu ja yhteinen dokumentointi on selkeyttänyt tilannetta ja ryhmä on sekä muotoillut periaatteelliset suositukset metatietojen tuottamiselle ja käsitteistön että laatinut ehdotuksia jatkotoimenpiteiksi. Alan kehitys on todella nopeaa ja olisi tärkeää olla aktiivinen sekä kansallisella tasolla että kansainvälisesti, jotta kehitys jatkuu oikeasuuntaisena.

Ryhmässä on hahmoteltu, miten työtä tulisi jatkaa metatietojen osalta, jotta ryhmän tavoite laadukkaista, koneluettavista käyttöoikeuksien hallintaan liittyvistä metatiedoista voidaan toteuttaa. Ryhmätyötä on pidetty hyvänä toimintamallina.

Tämä alustava selvitysteksti on syntynyt Kansallinen digitaalinen kirjasto- sekä Avoin tiede ja tutkimus -hankkeiden asiantuntijoiden yhteistyönä keväällä 2016. Laajan työryhmän tavoitteena on ollut edistää tietoaineistojen käyttöä ja avoimuutta laatimalla ehdotus aineistojen saatavuuteen, käyttöön ja niiden rajoituksiin liittyvistä metatiedoista. Metatietojen tulisi olla selkeitä, helppokäyttöisiä, yhteneviä ja koneluettavia. Tavoitteena on mahdollisimman hyvä semanttinen yhteentoimivuus, samalla kun on huomioitava eri toimijoiden hyvinkin erilaiset tarpeet, toimintakulttuurit ja tietojärjestelmät. Työ on pyritty asemoimaan siten, että se sopii sekä KDK- että ATT-hankkeiden kokonaisarkkitehtuurityöhön, ja huomioi kansainvälisen kehityksen.



Toinen luku toimii yleisenä johdantona aiheeseen ja luvussa kolme käymme läpi kontekstia ja kansainvälistä kehitystä.

Sähköisten aineistojen käyttö ja hallinnointi ovat ennestään ajankohtaistaneet oikeuksien hallinnan tarvetta tietotekniikan avulla. Tehokas palveluiden kehittäminen ja tiedonhallinnan parantaminen ovat oleellinen osa tieteen edistämistä ja yhteiskunnan digitalisoitumista. Sähköisiin aineistoihin ja niiden oikeuksien hallintaan liittyy monenlaisia haasteita. Osaltaan nämä juontavat puutteellisiin lisensseihin ja muihin sopimuksiin, metatietoihin ja vanhentuneisiin järjestelmiin, osittain myös arkaluontoisen tiedon, kuten henkilötietojen käsittelyyn, sekä siihen ettei tekijänoikeuslainsäädännössä riittävästi ole huomioitu tietoarkistoja ja sähköisen säilyttämisen ja tutkijoille käyttöön saattamisen vaatimaa joustavuutta.

Lukuun neljä olemme koonneet suosituksia parhaiksi käytännöiksi. Pyydämme lukijaa kiinnittämään huomiota varsinkin liitteisiin 1 ja 3, koska niistä ilmenee tarkempi tavoitetila.



Käsitteet

Aineisto

Yksilöitävissä oleva kokoelma sisältöjä

Käyttö

Toiminta jossa aineistoa käsitellään, esimerkiksi näytetään, kopioidaan, luetaan koneellisesti tai muokataan.

Käyttöoikeus

Valtuus/valtuutus käyttää aineistoa tai sen ilmentymää tai sen osaa. Vrt käyttö.

Käyttörajoitus

Ehto, joka liittyy aineiston käyttöön ja joka perustuu sopimukseen, lakiin tai etiikkaan, ja joka toteutuessaan voi estää pääsyn aineistoon tai määrittelee miten aineistoa saa käyttää.

Lisenssi

Kirjallinen käyttö lupa, johon voi liittyä rajoituksia. Lisenssi liittyy tiettyyn aineiston ilmentymään.

Oikeudenhaltija

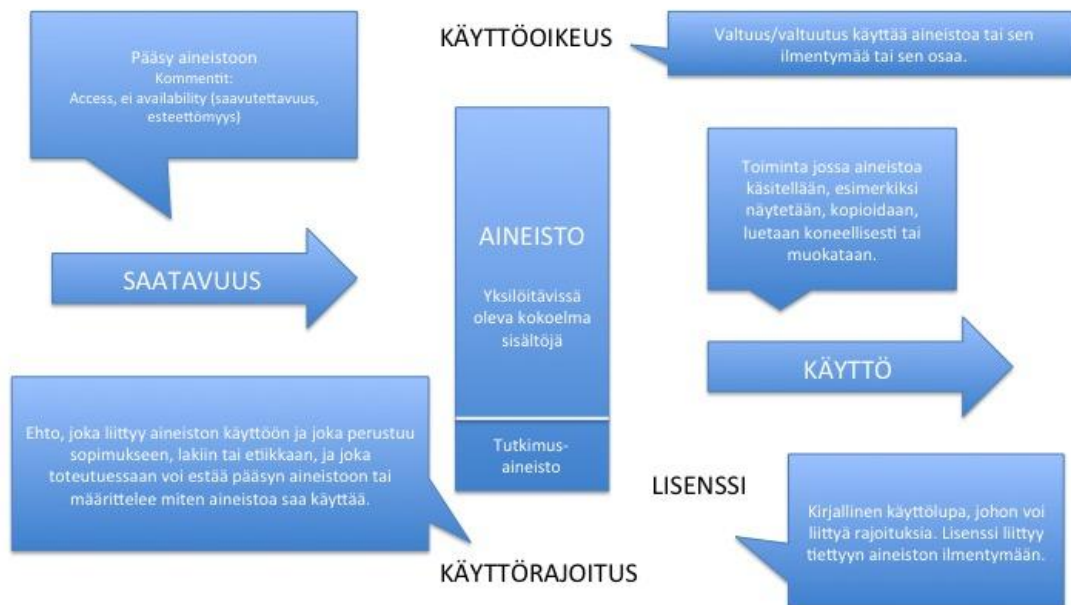
Luonnollinen henkilö, oikeushenkilö tai luonnollisten henkilöiden tai oikeushenkilöiden muodostama ryhmä, (ISO 27729) / toimija (JHS meta), joka voi päättää käyttöoikeuksien myöntämisestä.

Saatavuus

Pääsy aineistoon

Tutkimusaineisto

Tutkimuksessa tuotettava ja/tai käytettävä aineisto, jota analysoimalla pyritään vastaamaan asetettuun tutkimuskysymykseen.



Kuva 1 - Selvityksen keskeiset käsitteet

2. Oikeuksien hallinta ja metatiedot

Selkeät, dokumentoidut käyttöehdot ja -oikeudet ovat käytännössä ehdoton edellytys aineistojen levittämiselle ja jatkokäytölle. Myös löydettävyys on tärkeä osa käytettävyyttä ja tiedon avoimuutta. Sitä palvelevat erityisesti metatietojen aggregointiin perustuvat hakupalvelut kuten Finna ja Etsin. Jotta käytettävyys olisi hyvää, myös oikeuksien hallintaan liittyvien metatietojen tulisi olla mahdollisimman yhtenäisiä ja käyttäjän kannalta selkeitä. Siksi semanttinen yhteentoimivuus on tärkeää. Vaikka eri järjestelmissä ja toimintaympäristöissä voi olla hyvinkin erilaisia tarpeita käyttöoikeuksien metatietojen suhteen, voidaan yhtenäisistä kuvailukäytännöistä silti sopia. Voimme sopia ja ohjeistaa sen, mitä asioita metatiedoissa on syytä kertoa ja millä tavalla. Monet käyttöön ja saatavuuteen liittyvät rajoitukset perustuvat samoihin lainkohtiin, jolloin on mahdollista ilmaista asia samalla tavoin riippumatta käytetystä metatietoformaattista. Yhteismitallisuus loisi edellytykset tietojen tehokkaalle koneelliselle käsittelylle ja helpottaisi niiden tulkintaa muutenkin.

Perinteisesti käyttöoikeuksien metatietojen tallentamiseen ei ole panostettu, koska esimerkiksi painetun julkaisun käytössä ei ollut mitään epäselvää. Tarve näiden tietojen syöttämiseen syntyi vasta digitaalisten aineistojen myötä. Käyttäjien pitäisi voida helposti selvittää, mitä he saavat tehdä sähköisille aineistoille sen jälkeen kun he ovat kopioineet ne itselleen.

Julkaisujen, asiakirjojen ja muiden aineistojen takautuvat digitointihankkeet ja verkkojulkaiseminen ovat tehneet oikeuksienhallinnasta akuutin ongelman ja käyttöoikeuksien metatietojen best practice -käytäntöjä haetaan niin kirjastoissa, arkistoissa ja museoissa kuin tutkimussektorillakin. Tarvitaan sekä



kuvailusääntöjä että metatietoformaatteja/tietomalleja, jotka kattavat käyttöoikeuksien metatiedot. Ohjeistuksen on oltava tarkka ja formaattien ajan tasalla, jotta tuotetut metatiedot olisivat sekä ihmisluettavia että koneymmärrettäviä.

Yksinkertaisimmillaan käyttöoikeuksien metatiedoissa kerrotaan vain, että aineisto on vapaasti käytettävissä. Mutta jos aineisto ylittää teoskynnyksen ja jos tekijänoikeus on voimassa, tarvitaan jo enemmän tietoja. Jos tekijöitä on paljon eikä aineiston käytöstä ole solmittu lisenssiä, käyttöoikeustiedot voivat olla hyvin monimutkaisia. Oikeuksien haltijat tulisi luetella mahdollisuuksien mukaan kattavasti, mikä voi olla perikuntien osalta haasteellista. Myös ei-julkisten aineistojen käyttöoikeuksien kuvaaminen voi olla vaativaa.

Vaikka jokin aineisto olisi tekijänoikeuden kannalta vapaa, se ei välttämättä ole julkinen. Jos aineisto on salassa pidettävä, käyttöoikeuksien metatiedoissa tulisi kuvata salassapidon peruste, turvallisuusluokka, salassapitoaika ja päivämäärä jolloin aineisto vapautuu käyttöön sekä käyttöä koskevat poikkeussäädökset.

Karkeasti voidaan erottaa eri tyyppisiä tietoja, jotka liittyvät oikeuksien hallintaan ja joita voidaan liittää joko metatietoihin, aineistoihin tai niiden ilmentymiin tai osiin. Näitä ovat:

- Pääsyyn/saatavuuteen liittyvät ehdot
- Käyttöön liittyvät tiedot (mitä aineistolla saa tehdä)
- Kunkin rajoituksen peruste ja sen tyyppi
- Mahdollinen aikarajoitus
- Eri oikeuksien haltijoiden tiedot

Aineistojen avoimuutta ja käyttöä voivat rajoittaa monet asiat, kuten:

- Tekijänoikeus ja IPR
 - o tekijänoikeus
 - o tekijänoikeuden lähioikeudet (luettelot, tietokannat, teoskynnystä ylittämättömät valokuvat)
 - o patentit
 - o tavaramerkit
- Kansallista turvallisuutta koskevat lait
- Uhanalaisten lajien ja luontotyyppien suojelu
- Kulttuuriperinnön suojelu
- Alkuperäiskansojen suojelu
- Biopankkilait
- Tietosuojalait
- Tieteelliset embargosopimukset
- Luottamuksellisuus
- Yksityisyyden suoja
- Muut sopimukset
- Muut eettiset syyt

Käyttöoikeuksien metatiedoissa voidaan kuvata esimerkiksi:

- teoksen tekijänoikeudellinen tila
- tekijänoikeuden suojaaman aineiston oikeuksien haltija / haltijat, mahdollisuuksien mukaan kattavasti
- tekijänoikeuden suojaaman aineiston käyttöä koskevat lisenssit ja muut sopimukset kuten embargo aika



- aineistoa koskevat mahdolliset tietosuojamääräykset tai muut käytön rajoitteet

Käyttöoikeuksiin liittyviä metatietoja on kehitetty kahdella eri tavalla. Kirjastot, arkistot ja museot ovat lisänneet perinteisiin kuvailuformaatteihinsa kuten MARC 21:een käyttöoikeuksien kuvailun edellyttämiä tietoelementtejä. Tavoitteena on rikastaa perinteistä kuvailua riittävällä joskaan ei välttämättä kattavilla käyttöoikeuksien metatiedoilla. Näiden laajennusten takana ei ole ollut mitään hanketta tai selkeästi yksilöityä tarvetta, ja kokemusta käyttöoikeuksien metatietojen tallennuksesta näillä formaateilla on vielä verraten niukasti. Käyttöoikeuksien metatietoja kehitetään myös hankkeissa, kuten Europeanassa ja Euroopan komission johdolla DCat-ohjeistuksessa.¹ Getting it Right on Rights -hankkeessa haetaan kansainvälistä yhteisymmärrystä tekijänoikeuksien ilmaisemiseen liittyen sellaisille aineistoille, joille Creative Commons -käytäntö ei tällä hetkellä riitä. Tutkimusaineistojen käyttöoikeustietoja hahmotetaan esimerkiksi Research Data Alliancen ja CODATAN yhteisessä Legal Rights Interoperability -ryhmässä. Lisäksi myös W3C-konsortiolla on Permissions and Obligations Expressions -hanke, joka on käynnissä vuoden 2017 loppuun.

On vaikea arvioida, miten nämä lukuisat ja mahdollisesti keskenään ristiriitaiset hankkeet vaikuttavat esimerkiksi kirjastojen, arkistojen ja museoiden nykyisiin kuvailukäytänteisiin käyttöoikeuksien metatietojen osalta. Hankalasti ennakoitavia muutostekijöitä löytyy myös lainsäädännöstä. EU:n uusi tietosuojadirektiivi ja muut mahdolliset lakimuutokset, kuten tutkimuspoikkeukset, voivat saada seurauksia, jotka on huomioitava käyttöoikeuksien kuvailussa.

Selvityksessä on myös tuotu esiin monia muita tilanteita, joissa toimijoiden on vaikea toimia tarkoituksenmukaisesti nykyisen lainsäädännön, käytäntöjen ja järjestelmien puitteissa. Vaikka tässä selvityksessä keskitytään metatietoihin, muutkin huomiot ja tilannekuvaukset on otettu mukaan, sillä ne osaltaan kertovat tietyistä, olemassa olevista kehityspaineista ja toimijoiden toimintaympäristöstä.

On syytä varautua kansainvälisten ohjeiden muuttumiseen ja täsmentymiseen sekä mahdollisesti kansalliseen keskusteluun siitä, millaisia muutoksia tulisi tai voisi tehdä metatieto-ohjeistukseen ja esimerkiksi aineistojen tutkimuskäyttöön. On todennäköistä, että joudumme ottamaan huomioon muutoksia myöhemmin esimerkiksi rikastamalla tai konvertoimalla käyttöoikeuksien metatietoja. Tämä ei kuitenkaan millään tavalla ole syy olla käynnistämättä näiden tietojen tallennusta, sillä kaikki käyttöoikeuksien selventämiseksi tehty työ edistää suomalaisten aineistojen kansainvälistä näkyvyyttä ja käyttöä tulevaisuudessa.

Tärkeää on pyrkiä mahdollisimman hienojakoiseen käyttöoikeuksien kuvailuun ja tallennuksen riittävän täsmälliseen ohjeistukseen, sillä koneellisesti eri luokkien yhdisteleminen on huomattavasti helpompaa kuin olemassa olevien luokkien jakaminen jälkikäteen uusiin alaluokkiin. Jälkimmäinen vaatii valitettavan usein suuria määriä käsityötä. Siksi tässä dokumentissa annettavat suositukset (hyvät käytännöt) ovat osittain vaikeasti käytäntöön pantavia vanhojen aineistojen ja tietojärjestelmien kohdalla.^{2 3}

¹ DCat application profile https://joinup.ec.europa.eu/asset/dcat_application_profile/issue/licence-documents-and-rights-statements

² DCAT Application profile for data portals in Europe <https://joinup.ec.europa.eu/node/63567>

³ Permissions and Obligations <https://www.w3.org/2016/poe/charter>



2.1 Lisenssien kuvailu

Lisenssi on sopimus, jolla tekijä antaa jollekin käyttäjäryhmälle oikeuksia teoksensa hyödyntämiseen. Käyttöoikeuksien metatietoihin ei yleensä tallenneta lisenssiehtoja sellaisenaan, vaan esimerkiksi lisenssin nimi (kuten CC BY 4.0) sekä linkki verkossa olevaan sopimukseen tai sen tiivistelmään. Tiivistelmä voidaan luoda kuvailemalla lisenssi tähän tarkoitukseen kehitetyllä metadataformaattilla.

Lisenssien osalta kone- ja ihmisluettavuuden kärkeä edustavat Creative Commons -lisenssit. Niistä on verkossa tarjolla sekä kokoteksti että tiivistelmä monilla kielillä, myös suomeksi. Muitakin avoimia lisenssejä on useita, mutta nekin kattavat yleensä vain käyttöehtoja. Siksi yleisesti tunnetut lisenssit kattavat tällä hetkellä vain osan tarvittavista käyttöoikeuksiin liittyvistä asioista.

Lisenssiehtojen esittämiseen koneluettavassa muodossa on useita metadataformaatteja. Esimerkiksi Rights Expressions Language avulla voidaan kuvata lisenssejä XML- ja RDF-muodossa. Toinen samankaltainen väline on tieteellisten kustantajien kehittämä ONIX-PL eli Onix for Publications Licenses⁴, jonka avulla on mahdollista

- ilmaista lisenssit koneymmärrettävässä muodossa
- ladata ne sähköisiin sisällönhallintajärjestelmiin
- linkittää ne sähköisiin sisältöihin
- viestiä avainehdot käyttäjille
- verrata eri lisenssien ehtoja

Lisenssien ymmärtäminen voi olla haasteellista koneiden lisäksi myös ihmisille. Lisenssin kuvaaminen tähän tarkoitettuun formaattiin mahdollistaa lisenssiehtojen havainnollisen esittämisen. ONIX-PL -kuvauksen avulla lisensseistä voitaisiin tuottaa samankaltaisia tiivistelmiä kuin CC-lisensseistä on jo tehty manuaalisesti.⁵

Jos tutkimusdata asetetaan tarjolle jonkin CC-lisenssin mukaan, lisenssin itsensä kuvaamiseen ei ole erityistä tarvetta, koska linkki lisenssin verkossa olevaan tiivistelmään riittää. Mutta jos laaditaan paikallisia lisenssejä, on ratkaistava miten niiden ehdot voidaan esittää kompaktisti ja yksinkertaisesti jonkin tähän tarkoitettuun metadataformaatin avulla.

2.2. Pitkäaikaissäilytys ja käyttöoikeustiedot

Kansallisen digitaalisen kirjaston hankkeen PAS-palvelu toimii siten, että pitkäaikaissäilytettäviä aineistoja ei koskaan tarjota käyttäjille suoraan, vaan vain organisaatioiden omien taustajärjestelmien kautta. Toisin sanoen PAS-palvelun ei tarvitse tulkita normaaleja käyttöoikeuksien metatietoja, koska se lähettää aineistoa vain organisaatioiden taustajärjestelmiin tähän tarkoitukseen kehitettyjen rajapintojen kautta.

PAS-palvelu tarvitsee kuitenkin omat, PREMIS Rights -formaattiin perustuvat käyttöoikeuksien metatiedot, joilla määritellään kuka saa tehdä ja mitä

⁴ ONIX <http://www.editeur.org/21/onix-pl/>

⁵ Ks. CLARINin lisenssiehtojen määritelmät, joista voidaan automaattisesti generoida kokonainen loppukäyttäjän lisenssiteksti: <https://kitwiki.csc.fi/twiki/bin/view/FinCLARIN/ClarLegalDefinitions>.
esimerkki lisenssin instanssista: <https://kitwiki.csc.fi/twiki/bin/view/FinCLARIN/ClarEulaRes?PRIV=1>



pitkäaikaissäilytettävälle aineistolle. Kenties oleellisin raja on se, saako PAS-palvelun henkilökunta tehdä migraatioita eli muuntaa aineiston uuteen tiedostoformaattiin, vai rajataanko tämä oikeus aineiston tuottaja- ja omistajaorganisaation henkilökunnalle. Nämä linjaukset tulee tehdä aineiston siirtosopimuksessa. Aineiston tekijänoikeudellisella statuksella ei ole näiden linjausten kannalta mitään merkitystä, eikä esimerkiksi aineiston migraatio siirrä mitään oikeuksia sille taholle, joka sen on tehnyt.

On tietenkin tärkeää, että myös pitkäaikaissäilytettävien tutkimusaineistojen käyttöoikeudet kuvataan tarkasti ja yhteismitallisesti ja että nämä tiedot pidetään säilytyspaketeissakin ajan tasalla. Jos jostakin aineistosta vastannut organisaatio katoaa tai taustajärjestelmä josta aineisto oli hyödynnettävissä lakkaa toimimasta, viimeinen hätävara on PAS-palvelu ja siellä oleva kopio tutkimusdatasta metatietoineen. Aineiston asettaminen uudelleen tarjolle jostakin muusta taustajärjestelmästä ei ole mahdollista, elleivät metatiedot ole kaikin osin kattavat.

2.3 Rajapinnat eri järjestelmien välillä

Edellytyksenä metatietojen vaihtoon eri järjestelmien välillä on niiden semanttinen yhteentoimivuus. Käyttöoikeuksien metatietojen semanttinen yhteismitallisuus on ainakin toistaiseksi varsin rajallista. Myös syntaksin tasolla on haasteita, koska rakenteidenkin olisi oltava sellaisia että ne ovat yhdisteltävissä.

Yksittäisten tietoelementtien tasolla yhteismitallisuus edellyttää, että kuvailun periaatteiden yhtenäistämistä. Esimerkiksi samasta lisenssistä tulisi aina käyttää samaa nimeä tai lyhennettä (jotka ideaalitapauksessa poimitaan kontrolloidusta sanastosta, jossa termillä pitäisi olla myös tunniste). Formaatin tasolla yhteismitallisuus merkitsee sitä, että metatiedot ovat muunnettavissa formaatista toiseen, mikä edellyttää tietoelementtien yhteismitallisuutta. Kun sekä tietosisällöt että tietoelementit ovat yhtenevät ja formaattien tietoelementtien vastaavuudet ovat riittävät, metatietojen yhteismitallisuus on riittävä.

Eri formaattien tietoelementtien yhdistelemiseen tarjoavat apua metadatarokisterit, jotka linkittävät formaattien tietoelementit toisiinsa.⁶

Yleisenä metatietojen poiminnan rajapintamallina sovelletaan KDK-hankkeessa ja laajemminkin kirjasto- ja arkistoaloilla OAI-PMH -protokollaa.⁷ Jos tiedot siirretään suppeassa Dublin Core -perusmuodossa on käyttöoikeustiedoille käytettävissä vain Rights-elementti, eikä senkään soveltamiselle ei ole tarkempaa ohjeistusta. Dublin Core -formaattissa on kuitenkin muitakin käyttöoikeuksiin liittyviä kenttiä, joiden avulla voidaan tuottaa kohtuulliset käyttöoikeuksien metatiedot. Niiden tallentaminen pitäisi kuitenkin ohjeistaa tarkoin.

Julkaisuarkistojen osalta ATT-hankkeen rahoittaman TAJUA-projektin yhteistyöverkosto kehittää parhaillaan kansallista Dublin Core -pohjaista metadatasuosittelua, jonka odotetaan valmistuvan vuoden 2017 alkupuolella.⁸ Suosituksessa pyritään luomaan luomaan yhtenäiset, mahdollisimman yksinkertaiset käytännöt myös käyttöoikeuksien ilmaisemiselle julkaisuarkistojen

⁶ Ks esimerkiksi <http://iow.csc.fi/#/>

⁷ The Open Archives Initiative Protocol for Metadata Harvesting <http://www.openarchives.org/OAI/openarchivesprotocol.html>

⁸ Ks. TAJUA-yhteistyöverkosto <https://www.kiwi.fi/x/f4OFAw>



sisäisessä metadataformaattissa. Tavoitteena on, että julkaisuarkistoista OAI-PMH-rajapinnan ulospäin välitettäviin metadataformaatteihin voidaan tallennettujen tietojen pohjalta generoida tarpeen mukaan eri tavoilla muotoiltuja tietoja.

Finnaan⁹ kerätään aineistoja erittäin laajasti koko sektorilta ja siksi Finnan asettamilla vaatimuksilla on erittäin suuri vaikutus tuotantojärjestelmiin. Muita tietokantoja, joihin metatietoja kerätään OAI-PMH:lla tai muilla rajapintastandardeilla kuten Z39.50:llä ja SRU:lla ovat ainakin kirjastojen, museoiden ja arkistojen yhteisluettelot, OKM:n tiedonkeruu ja sen myötä Juuli. Kirjastojen MARC 21 -formaattilla voidaan Dublin Coren tavoin tallentaa varsin kattavat käyttöoikeustiedot, ja kirjastosektori on laatinut tallennuksesta seikkaperäiset ohjeet, mutta ne soveltuvat ensi sijassa vain julkaisuille.

Kansainvälisistä palveluista on hyvä mainita Europeana, jonne metadata viedään suomalaisista organisaatioista pääasiassa Kansalliskirjaston Formula-palvelun kautta, joka toimii Kansallisen digitaalisen kirjaston osana.

PAS-palveluiden kehittyessä aineistojen siirtäminen pitkäaikais säilytykseen edellyttää yhteismitallisia ja selkeitä käyttöoikeuksien hallintaan liittyviä metatietoja jotta palvelu toimisi tiedonhallinnan kannalta mielekkäällä tavalla. Aineistoja, joihin liittyvistä vastuista ja oikeuksista ei ole tarkkaa tietoa on ongelmallista säilyttää, koska niitä on vaikea tai mahdotonta saattaa käyttöön. Myös Kansallinen palveluväylä voi asettaa vaatimuksia, joita ei vielä tarkkaan tunneta, mutta joihin on syytä varautua tallentamalla aineistoille riittävän kattavat metatiedot.

⁹ Aineistojen haravointi Finnaan Finna-wikissä <https://www.kiwi.fi/x/PA0XAw>



3. Käyttöoikeuksien kuvaaminen eri metadastandardeissa

Käyttöoikeuksiin ja niiden esittämiseen liittyvät kysymykset ovat kansainvälisestikin erittäin ajankohtaisia ja asiaa pohditaan monissa työryhmissä ja hankkeissa. Onkin todennäköistä, että uusia lisenssejä, ohjeistuksia, käytäntöjä ja myöhemmin myös uusia tai päivitettyjä metadastandardeja on tulossa lähivuosina.

Suurimmat haasteet eivät ole teknisiä, vaan hallinnollisia ja epäselvistä käytännöistä ja olosuhteista johtuvia. Aineistoihin liittyviä kirjallisia käyttösopimuksia ei aina ole tehty lainkaan tai ne ovat jatkokäytön kannalta liian yleisiä, eikä lainsäädäntökään tarjoa helppoja ratkaisuja kaikissa ongelmatilanteissa. Jos lähtötilanne on tämä, edellytykset metatietojen tallentamiseen eivät ole kaksiset.

Perinteisten julkaisujen käyttöoikeuksien osalta tilannetta kuvaa hyvin seuraava kustannusalalla pitkään töitä tehneen henkilön kommentti:

I'm a bit ambivalent about the rights metadata because of the total lack of progress. Technically the solution is there. There are schemas ready to use etc. I'm disappointed that the rights owners / publishers aren't supporting this and investing. The trouble is that on the one hand they rely on blanket licenses which is easy money for them and on the other they are scared of their image as rights enforcers.

And the ugly truth is that their back office systems are just not fit for purpose.

Kustantajien sovellusten tavoin myöskään kirjastojen ja muiden muistiorganisaatioiden taustajärjestelmät eivät osaa hyödyntää käyttöoikeuksien metatietoja kattavasti. Toisaalta esimerkiksi MARC 21 -formaatti ei sovellu järin hyvin esimerkiksi Suomen lainsäädäntöä vastaavien käyttöoikeuksien metatietojen tallennukseen, koska formaatin nykyinen versio perustuu amerikkalaiseen lainsäädäntöön. Suomalaisia tietoja varten formaattiin tarvitaan muutamia laajennuksia.

Pyrimme selvityksen tässä osassa kartoittamaan meneillään olevaa kehitystä sekä jo käytössä olevia tietomalleja ja käytäntöjä, jotta pystymme tekemään mahdollisimman kestäviä suosituksia. Selvityksen liitteessä 1 on esitetty esimerkkejä, siitä miten oikeuksien hallintaan liittyviä metatietoja esitetään useissa eri formaateissa.

3.1 Kansainvälinen yleinen kehitys

DCAT

DCAT on Dublin Coreen perustuva, W3C-konsortion ylläpitämä datakatalogien kuvailuun luotu RDF-sanasto ja skeema, josta Euroopan unioni on tehnyt oman sovellusprofiilin, jota edelleen kehitetään, kokemuksen karttuessa julkisen datan avaamisessa. Se on tehty virastotyötä varten PSI-direktiivin toteuttamiseen ja avoimen datan hallinnoinnin helpottamiseksi ja löydettävyyden parantamiseksi.

Siitä on useita sovellusprofiileja ja niitä kehitetään esimerkiksi geotalle ja tilastoille. Myös ATT-hankkeessa on huomioitu DCAT-yhteensopivuus. DCATIin liittyy myös W3C:n suosittelema sanasto.¹⁰

DCAT on siis keskeinen, mutta myös vasta kehittyvä ohje. Käyttöoikeuksien metatietojen tallennusta varten standardissa käytetään Dublin Core rightsStatements-elementtiä kuvaamaan tutkimusaineistoon liittyviä tekijänoikeuksien tai pääsyn ehtoja. Lisäksi tutkimusaineiston jakeluversiolle (Distribution) voidaan määritellä lisenssi Dublin Coren License-elementillä. Lisenssien sanastoiksi suositellaan Creative Commons lisenssiä, Open Data Commons Public Domain Dedication and Licenseä (PDDL), ISA Open metadatalisenssiä, EU:n EUPL tai jäsenvaltioiden omia avoimia lisenssejä. Lisenssejä kehitellään edelleen Open Data Instituten Open Data Rights Statement Vocabularyssa¹¹ sekä Open Digital Rights Languageissa (ORDL).¹² Myös lisenssien luokittelua ollaan tekemässä.¹³ Aikaisemmin luotu eurooppalaisten info:eu-repo - yhteentoimivuustyö on loppunut, mutta EU:ssä käytetään edelleen sen luomia saatavuusluokituksia tai niiden muunnelmia.¹⁴

Europeana

Europeanan ja DPLA:n Rightsstatements.org -yhteistyö on ehdottanut seuraavia luokkia tekijänoikeuksiin liittyvien oikeuksien suhteen (White paper updated in January 2016).¹⁵ Marraskuussa 2016 Europeana otti käyttöön näistä kuusi eri arvoa (keltaisella pohjalla):¹⁶

In Copyright (inC)	Tekijänoikeuden alainen.
In Copyright EU Orphan Work (InCOWEU)	Orpoteos EU:ssa.
In Copyright Rightsholder(s) Unlocatable or Unidentifiable (InCRUU)	Tekijänoikeuden alainen, mutta tekijää ei ole löydetty tai tavoitettu.
In Copyright Educational Use Permitted (InCEDU)	Tekijänoikeuden alainen teos, opetuskäyttö sallittu ilma erillistä lupaa.
In Copyright NonCommercial Use Permitted (InCNC)	Tekijänoikeuden alainen teos, vain ei-kaupallinen käyttö sallittu ilman erillistä lupaa.

¹⁰ DCat vocabulay <https://www.w3.org/TR/vocab-dcat/>

¹¹ Open Data Institute. Open Data Rights Statement Vocabulary. <http://schema.theodi.org/odrs/>

¹² Open Data Commons Public Domain Dedication and License (PDDL) <http://opendatacommons.org/licenses/pddl/>, ISA Open Metadata Licence v1.1, <https://joinup.ec.europa.eu/category/licence/isa-open-metadata-licence-v11>, European Commission. Joinup. Open Source Software. European Union Public Licence (EUPL). <http://joinup.ec.europa.eu/software/page/eupl>.

¹³ https://joinup.ec.europa.eu/asset/dcat_application_profile/description ; <https://www.w3.org/TR/vocab-dcat/>

¹⁴ Info-eu-repo, Wiki <https://wiki.surfnet.nl/display/standards/info-eu-repo#info-eu-repo-AccessRights>]

¹⁵ Rightsstatements.org White Paper http://rightsstatements.org/files/160208recommendations_for_standardized_international_rights_statements_v1.1.pdf

¹⁶ Europeana Factsheet pro.europeana.eu/files/Europeana_Professional/IPR/rightsstatement.org/europeanafactsheet.pdf

No Copyright NonCommercial Use Only (NoCNC)	Ei voimassa olevaa tekijänoikeutta, vain ei-kaupallinen käyttö sallittu.
No Copyright Contractual Restrictions (NoCCR)	Ei voimassa olevaa tekijänoikeutta, mutta sopimukseen perustuvia rajoituksia.
No Copyright Other Known Legal Restrictions (NoCOKLR)	Ei voimassa olevaa tekijänoikeutta, mutta muita lakiin perustuvia rajoituksia jatkokäytön suhteen.
No Copyright United States (NoCUS)	Ei USA:ssa voimassa olevaa tekijänoikeutta.
No Known Copyright (NKC)	Ei tunnettua tekijänoikeutta.
Copyright Not Evaluated (CNE)	Tekijänoikeuksia ei ole selvitetty.

Taulukko kertoo omalla tavallaan käyttöoikeustietojen monimutkaisuudesta. Jos tekijänoikeus ei ole voimassa, on vaikea nähdä kenellä olisi oikeus rajoittaa aineiston käyttöä sopimuksin tai vaatia kaupallisen käytön kieltoa. Lisäksi eri maita koskevien tekijänoikeuden rajoitusten merkitsemisestä tulee hyvin hankalaa jos pyritään kattavuuteen.

On tärkeä huomata, että nämä tekijänoikeuksiin liittyvät metatiedot eivät korvaa pääsyyn liittyviä metatietoja, sillä tekijänoikeudet voivat koskea samalla tavalla avoimia kuin suljettujakin aineistoja. Ne eivät siten liity siihen, miten aineistoja saa käyttöönsä, vaan siihen, mitä niillä voi tehdä, kun ne kerran on saanut. Sen tähden myös erillinen tieto pääsystä on tarpeen.¹⁷

RDA ja CODATA

Research Data Alliance -verkosto on perustanut CODATAN kanssa yhteisen Legal Interoperability -intressiryhmän (Interest Group) jossa pyritään edistämään kansainvälistä yhteentoimivuutta tutkimusaineistojen oikeuksien hallinnassa.¹⁸ Ryhmä laatii parhaillaan avoimuutta tukevia periaatteita ja hyviä käytäntöjä, sillä siten voidaan edistää myös yhteensopivuutta ja datan uusiokäyttöä. Kun käsitellään sähköisiä aineistoja on hyödyllisempää lähestyä aihetta vastuiden ja oikeuksien kautta kuin keskustella omistajuudesta. Tämä poikkeaa esimerkiksi JISCin ohjeista, joissa on perinteisesti painotettu nimenomaan tutkimusdatan omistajuuden määrittelyn tärkeyttä. On mahdollista, että tässäkin suhteessa kehittyä omanlaisensa anglosaksinen lähestymistapa, mutta voi myös olla, että tämä paperimaailman ajattelumalli on väistymässä.

RDA:n mukaan tutkijoiden vastuulla on selkeästi kertoa millä oikeuksilla aineistoihin pääsee käsiksi ja miten niitä saa käyttää (access and reuse). Tutkijoiden tulisi välttää erillisiä sopimuksia, joissa luodaan rajoituksia, sekä varsinkin tekijänoikeuksien vaatimista, jos niiden syntymistä voi välttää. Data-

¹⁷ Rightsstatements.org White paper.

http://rightsstatements.org/files/160208recommendations_for_standardized_international_rights_statements_v1.1.pdf ; Getting it Rights on Rights <http://dp.la/info/about/projects/getting-it-right-on-rights/>

¹⁸ Legal Interoperability Interest Group RDA <https://rd-alliance.org/groups/rdacodata-legal-interoperability-ig.html>



arkistojen tulisi huolehtia oikeuksien hallintaan liittyvistä metatiedoista. Niihin tulisi aina sisältyä vähintään tiedot avoimuuden asteesta sekä käyttöehdoista. Viranomaiset ja tutkimusorganisaatioiden tulee tehdä datapolitiikat ja edistää niissä avoimuutta ja perustella mahdolliset rajoitukset lainsäädännöllä.

Europeana Data Model

Europeanaan kulttuuriperintöaineistoa toimittavilla kirjastoilla, arkistoilla ja museoilla on käytössään erilaisia metadataformaatteja, kuten Dublin Core, EAD ja LIDO. Europeana Data Model (EDM) on suunniteltu mukauttamaan eri formaatteista saatu kuvailutieto sellaiseen muotoon, että sitä voidaan esittää ja hyödyntää erilaisten aineistojen kannalta tarkoituksenmukaisesti European kaltaisessa kansainvälisessä, metadataa laajasti aggregoivassa kontekstissa. Tarkoituksena on, että alkuperäisen datan rikkaus pystytään säilyttämään mahdollisimman hyvin. EDM on suunniteltu myös semanttisen webin ja linkitetyn datan mahdollisuudet huomioiden. EDM:ää edelsi European Semantic Elements (ESE). EDM on yhteensopiva ESE:n kanssa ja Europeana tukee edelleen myös ESE-muotoista dataa.¹⁹

Formula on Kansalliskirjaston ylläpitämä palvelu, johon poimitaan metatietoja Formula-sopimuksen allekirjoittaneiden tahojen tietokannoista, ja josta nämä tiedot ovat haravoitavissa Europeanaan. Konversiosäännöt organisaation käyttämästä formaatista EDM:ään laatii metatiedon luovuttava organisaatio.²⁰

DataCite ja NISO

DataCite on tutkimusaineistojen hallintaan keskittynyt yhteisö. Sen oma metadataskeema ei tunne pääsyn rajoituksia kuvaavaa metatietoa. RightsList-alueelle voi kuitenkin tallentaa oikeuksienhallintaan ja käyttöehtoihin liittyvää tietoa. Kenttään voidaan tallentaa esimerkiksi lisenssin nimi ja URL-osoite. Lisäksi on käytettävissä esim. Contributor-tietoelementin RightsHolder-tietoelementti, johon voidaan tallentaa oikeuksien omistaja.²¹

Yhdysvalloissa ANSI/NISO (American National Standards Institute / National Information Standards Organization) on tuottanut Access License and Indicators -suosituksen²². Se suosittelee käyttämään free_to_read elementtiä sen kuvaamiseen että aineisto on vapaasti luettavissa, ja license_reference-elementtiä sen kertomiseen, miten aineistoa saa käyttää. Kenttään voidaan tallentaa esimerkiksi linkki verkossa olevaan lisenssiin.. free_to_read -elementtiin voidaan lisätä alku- ja loppupäivämäärät. Lisenssin muuttuminen tiettyinä ajankohtana voidaan myös kuvata metatiedoissa.

NISO:n ohje sisältää erilaisia käyttötapauksia joihin lisenssien tulisi vastata. Esimerkki tekstin louhimisesta:

Description: Currently it is difficult and labor-intensive for text miners to know what they can access for mining content (which articles and which parts of articles, for example), and what they can do with the information they mine.
[---]

¹⁹ Lisätietoa <http://pro.europeana.eu/page/edm-documentation>

²⁰ Formula Kansallisen digitaalisen kirjaston sivuilla <http://www.kdk.fi/index.php/fi/europeana-ja-muut-hankkeet/formula>

²¹ DataCite Metadata Kernel http://schema.datacite.org/meta/kernel-3/doc/DataCite-MetadataKernel_v3.1.pdf

²² Access License and Indicators. National Information Standards Organization. NISO RP-22-2015 http://www.niso.org/apps/group_public/download.php/14226/rp-22-2015_ALI.pdf

Publishers should provide a machine-readable <license_ref> URI in the article metadata that references any part of the article licensed under a different license to that governing the article as a whole. Licenses should indicate whether or not text and data mining is allowed and whether there are additional terms and conditions. Note that interpretation of re-use rights would be the responsibility of the user.

NIOS:n ohjeistukseen sisältyy myös esimerkkejä XML-skeemoista joilla suosituksen mukaiset metatiedot saadaan koneluettaviksi.



4. Suositukset oikeuksien hallintaan liittyvien metatietojen luomiseen ja käyttöön

Selkeät, dokumentoidut käyttöehdot ja aineistoihin liittyvät oikeudet ovat käytännössä ehdoton edellytys sähköisten aineistojen levittämiseksi ja jatkokäytölle. Löydettävyyttä ja laadukkaita metatietoja ovat tärkeä osa tiedon käytettävyyttä ja avoimuutta. Avoin tiede ja tutkimus ja Kansallinen digitaalinen kirjasto -hankkeissa on luotu yhteiset suositukset hyviksi käytännöiksi siitä, miten käyttö- ja saatavuusehtoihin liittyvät tiedot tulisi dokumentoida aineistojen metatietoihin. Alla esitetyt ohjeet ovat suosituksia, joita noudattamalla sähköisten aineistojen käyttöä ja leviämistä voidaan edistää.

1. Avoimuuteen ja kestävään käyttöoikeuksien hallintaan on pyrittävä aineiston koko elinkaaren ajan. Ellei aineiston tai metadatan käyttöä tai saatavuutta ole rajoitettava lain, sopimuksen tai eettisen syyn perusteella sen tulisi olla avointa.
2. Sovella käyttöoikeuksien kuvailussa suositusten mukaisia standardeja ja ohjeita. Erotta käyttöä ja saatavuutta (pääsyä) koskevat rajoitukset toisistaan.
3. Kaikki rajoitukset on kuvattava metatiedoissa sekä ihmisettä (jos mahdollista) koneluettavalla ja -ymmärrettävällä tavalla.
4. Käyttörajoitusten tarkan kohteen tulee ilmetä yksiselitteisesti ja selvästi (esim. metatiedot, koko kokoelma, aineisto tai jokin yksittäinen tieto metatiedoissa tai aineistossa tai aineiston tietty ilmentymä).
5. Jos aineisto ei ole avoimesti saatavana ja käytettävänä tulee perusteisiin viitata (esim. laki tai sopimus), mielellään sekä ihmisettä koneluettavalla tavalla.
6. Tekijänoikeuksiin perustuvat käyttörajoitukset tulee, jos mahdollista, kertoa Creative Commons -lisenssien avulla. Jos ne eivät sovellu, voidaan käyttää muita laajalti käytettyjä ratkaisuja, kuten Europeanan ja DPLA:n rightsstatements -suosituksia. Metatietoihin tulisi tallentaa lisenssin nimi tai sen lyhenne sekä - jos mahdollista - linkki lisenssiin.
7. Mikäli käytössä on organisaatio- tai aineistokohtaisia käyttörajoituksia, niiden noudattamisen tulee olla helppoa aineistojen käyttäjille. Kuvaile selkeästi, mitä käyttäjä saa tehdä aineistoille ja tallenna linkki metatietoihin. Kehitä ja ota käyttöön palveluja, jotka tukevat aineistojen haettavuutta käyttörajoitteiden perusteella ja järjestämistä käyttörajoitusten mukaan.
8. Oikeuksien haltijat on ilmaistava mahdollisimman yksiselitteisesti.
9. Aikarajoitukset ja päiväykset on huomioitava metatiedoissa ja ne on kerrottava yksiselitteisesti ja standardien mukaisesti.