

Oikeuksien hallintaan liittyvät metatiedot -selvitys

Versio 0.95

Sisällys

1. Johdanto

Käsitteet

2. Oikeuksien hallinta ja metatiedot

Lisenssien kuvailu

Pitkäaikaissäilytys ja käyttöoikeustiedot

3. Suomen kulttuuriperinnön ja tieteen tietoinfrastruktuurit

Arkistolaitos ja arkistot

Yksityisarkistot

Kirjastot

Museot

Kansallinen digitaalinen kirjasto ja Finna

Tutkimusaineistojen arkkitehtuurit

Tutkimusorganisaatiot

Tieteenalariippumattomat tutkimusaineistopalvelut

Ihmistieteellisten alojen tutkimusaineistot

Sensitiiviset henkilötietoaineistot

Luonnonvara- ja ympäristöaineistot

Esimerkkejä tutkimusaineistoista

Pitkäaikaissaatavuus


Rajapinnat eri järjestelmien välillä

4. Käyttöoikeuksien kuvaaminen eri metadastandardeissa

Kansainvälinen yleinen kehitys

5. Suositukset oikeuksien hallintaan liittyvien metatietojen luomiseen ja käyttöön

6. Etenemisehdotus

Julkaisu Oikeuksien hallintaan liittyvät metatiedot -selvitys	
Julkaisija Avoin tiede ja tutkimus -hanke	Julkaisuajankohta xx.2016
Tekijä Avoin tiede ja tutkimus -hanke	
Lisenssi  Tämä teos on lisensoitu Creative Commons Nimeä 4.0 Kansainvälinen -lisenssillä .	
Julkaisun jakelu avointiede.fi/	
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 Hupaniitty, Suomalaisen Kirjallisuuden Seura
Nina Hyvönen, 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, 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 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. Kolmannessa luvussa kuvaillaan erilaisia toimijoita eli "arkkitehtuureja", toimintaympäristöjä ja järjestelmiä, joita olemme tunnistanee. Olemme pyrkineet kuvaamaan niiden tämänhetkistä tilaa ja tarpeita oikeuksien hallintaan liittyen, sekä tiedonhallinnallisessa mielessä että käytännön tasolla. Koska muutosvauhti on kova, olemme halunneet antaa aiheelle tilaa ja koota yleiskuvan. Toivomme, että prosessista ja lopputuloksesta on apua kaikille, jotka toimivat aineistojen kanssa.

Luvussa neljä 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 viisi olemme koonneet suosituksia parhaiksi käytännöiksi. Lukuun kuusi olemme kirjanneet ehdotuksia jatkotoimenpiteiksi. 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

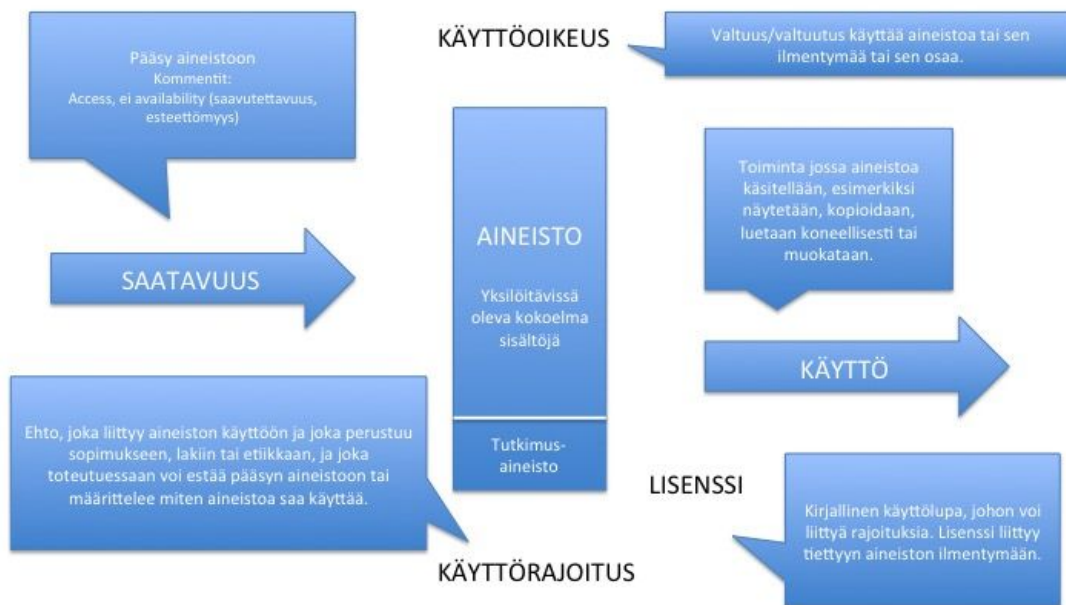
Luonnollinen henkilö, oikeushenkilö tai luonnollisten henkilöiden tai oikeushenkilöiden muodostama ryhmä, (ISO 27729) / toimija (JHS meta), joka voi päättää käyttöoikeuksien myöntämisestä.

Saatavuus

Pääsy aineistoon

Tutkimusaineisto

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



Kuva 1 - Selvityksen keskeiset käsitteet

2. Oikeuksien hallinta ja metatiedot

Selkeät, dokumentoidut käyttöehdot ja -oikeudet ovat käytännössä ehdoton edellytys aineistojen levittämiseksi 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 riittäväillä 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}

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 ihmislueuttavuuden 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.

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>

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 tarkoitettun metadataformaatin avulla.

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 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.

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>

3. Suomen kulttuuriperinnön ja tieteen tietoinfrastruktuurit

Kirjastot, arkistot ja museot ovat pitkän linjan toimijoita aineistohallinnan alalla. Erilaiset toiminnat, tarpeet ja aineistot ovat kuitenkin synnyttäneet eroja niiden käytännöissä vuosikymmenten ja -satojen mittaan. Kansallinen digitaalinen kirjasto –hankkeessa on kuitenkin jo useita vuosia ratkottu näitä käyttäjälle sähköisessä ympäristössä hankalia ongelmia. Mukana työssä ovat olleet myös esimerkiksi Yhteiskuntatieteellinen tietoaarkisto ja KAVI. Kenttä on hyvin laaja ja monimuotoinen. Osittain tiedonhallintaa on kyetty ohjaamaan keskitetysti, mutta organisaatioissa on hyvinkin erilaisia tarpeita ja resursseja. Kirjastoissa tietojärjestelmät ovat olleet sähköisiä jo pitkään ja metatietoja on tuotettu suhteellisen yhteismitallisesti. Sen sijaan arkistoissa ja museoissa aineistot ovat usein vanhoja ja siksi aineistojen moninaisuus ja oikeuksien kompleksisuus voivat olla todella haastavia, kun tavoitteena on kuvailevan metatiedon yhteismitallisuus ja rikkaus. On myös hyvä muistaa, että kulttuuriperintöaineistot ovat tutkijoiden tärkeitä lähteitä.

Arkistolaitos ja arkistot

Viranomaisarkistoja ohjaavat lait ja säädökset, kun taas yksityisarkistoja voivat koskea myös sopimukset. Kun yksityinen arkistonmuodostaja luovuttaa arkistoaineistonsa arkistolaitokseen, laaditaan aineiston vastaanoton yhteydessä vastaanottotodistus, jonka allekirjoittavat sekä arkistoaineiston luovuttaja että aineiston vastaanottava arkistolaitoksen edustaja. Vastaanottotodistuksen mukaan aineisto luovutetaan ja lahjoitetaan arkistolaitokseen vapaasti käytettäväksi. Mikäli yksityinen arkistonmuodostaja haluaa rajoittaa luovuttamansa ja lahjoittamansa arkistoaineiston käyttöä, laaditaan vastaanottotodistuksen lisäksi erillinen luovutus sopimus. Luovutus sopimuksessa arkistonmuodostaja voi esittää, millä tavoin hän haluaisi arkistoaineiston käyttöä rajoittaa. Arkistoaineiston vastaanottaja ja luovuttaja sopivat yhdessä käyttörajoitusten laajuudesta. Nämä sopimukset eivät ole "määrämuotoisia" eivätkä standardoituja, joskin tähän suuntaan pyritään. Myös sopimuksia voisi kuitenkin manuaalisesti luokitella ja kuvailla metatietoihin.

Yksityisarkistoon luovutettua arkistoaineistoa kutsutaan yleensä N.N.:n arkistoksi joko luovuttajan nimen mukaan tai sen henkilön nimen mukaan tai sen yhteisön mukaan, jota arkistoaineisto koskee. Yksityisarkistojen (N.N.:n arkisto) lisäksi on olemassa myös kokoelmia. Kokoelma on eri asia kuin yksityisarkisto. Kokoelma on syntynyt määrätietoisesta keräilytyön tuloksena. Arkistolaitoksen Vakka-tietokannan mukaan Kansallisarkistossa (Helsinki) on noin 915 yksityisarkistoa, joissa on jokin käyttörajoitus. Kaiken kaikkiaan Kansallisarkistossa on yksityisarkistoja sijoitusrekisterin mukaan noin 2 900, joten käyttörajoitettuja yksityisarkistoja Kansallisarkistossa on noin 31 % kaikista Kansallisarkiston yksityisarkistoista. Nykyään arkistolaitoksessa pyritään siihen, että yksityisarkistojen käyttö olisi mahdollisimman vapaata. Jos käyttörajoituksia asetetaan, niiden tulisi olla hyvin perusteltuja, selkeitä ja yksinkertaisia. Yksityisarkistoille asetettujen käyttörajoitusten määrä onkin laskenut voimakkaasti. Myös kokoelmiin voi kohdistua käyttörajoituksia.

Organisaatioiden arkistonmuodostus on usein erilaista kuin yksityisten ihmisten. Yksityisten yhteisöjen arkistointi- ja asiakirjahallinto voi vaihdella huomattavasti. Ainakin viranomaisten asiakirjahallinnon tietojärjestelmien taustalla on sähköinen arkistonmuodostussuunnitelma (eAMS). Arkistonmuodostussuunnitelma on organisaation tekemä suunnitelma tehtävien hoidon tuloksena kertyvien asiakirjojen säilytysajoista ja -tavoista ja julkisuus- ja henkilötietolain mukaisista

käyttörajoituksista. Arkistonmuodostussuunnitelmassa on muutakin asiakirjojen ja tietojen säilytykseen, käsittelyyn ja käyttöön liittyvää informaatiota.

Sähköisiä asianhallintajärjestelmiä käyttävien viranomaisten on käytettävä arkistolaitoksen Sähke2-määräyksen mukaan toimivia tietojärjestelmiä. Sähke2-määräys koskee asiakirjallisten tietojen käsittelyä, hallintaa ja säilyttämistä. Sähke2-määräys antaa viranomaisille ja muille määräyksen mukaan toimiville organisaatioille metatietomallin, jonka mukaan asiakirjallinen tieto kuvataan.

Tämän arkistoaineistojen käyttörajoituksia koskevan koosteen kannalta keskeistä Sähke2:ssa ovat käyttörajoitus-metatietoelementin tarkentavat metatietoelementit:

- *Julkisuusluokka*, jonka arvot ovat julkinen, osittain salassa pidettävä, salassa pidettävä
 - o Jos ei ole julkinen, niin on kerrottava
 - salassapitoaika, jonka arvo on kohteen salassapitoaika vuosina
 - salassapidon päättymisajankohta, jonka arvoksi tallennetaan se hetki, jolloin tieto muuttuu julkiseksi
 - salassapitoperuste, jonka arvo on tieto siitä, mihin salassapito perustuu
- *Suojaustaso*, jonka arvo on joko I, II, III tai IV
- *Turvallisuusluokka*, jonka arvo on joko 1, 2, 3, 4 tai "ei turvallisuusluokiteltu"
- *Käyttäjryhmä*, jonka tietoelementit ovat henkilö, rooli ja kuvaus
 - o Henkilö-tietoelementin arvo on tieto henkilöstä, jolla on tehtävässään oikeus käsitellä salassa pidettäviä aineistoja. Oikeudet voidaan määritellä henkilöittäin ja/tai rooleittain
 - o Rooli-tietoelementin arvo on tieto roolista, jolla on tehtävässään oikeus käsitellä salassa pidettäviä aineistoja
 - o Kuvaus- tietoelementin arvo on kuvaus toimenpiteistä, joihin henkilöllä/roolilla on oikeus (1 = lukuoikeus, 2 = muokkaus-oikeus)

Arkistolaitoksen sopimukset

Nykyään arkistolaitoksessa pyritään siihen, että yksityisarkistojen käyttö olisi mahdollisimman vapaata. Jos käyttörajoituksia asetetaan, niiden tulisi olla hyvin perusteltuja, selkeitä ja yksinkertaisia. Yksityisarkistoille asetettujen käyttörajoitusten määrä onkin laskenut voimakkaasti. Yleisimmät yksityisarkistojen käyttörajoitustyyppit arkistolaitoksessa ovat:

1. Arkiston käyttö valtionarkistonhoitajan/arkistolaitoksen pääjohtajan/maakunta-arkiston johtajan luvalla.
2. Arkiston käyttö arkistoaineiston luovuttajan luvalla. Arkiston luovuttaja voi olla yksityinen henkilö, perhe, suku, yhdistys, säätiö tai muu vastaava.
3. Arkiston käyttörajoitus on 1- ja 2-kohdan yhdistelmä. Arkiston käyttöön edellytetään lupa arkistolaitoksen johtajalta ja luovuttajalta.
4. Arkiston käyttörajoitus on jokin muu. Arkiston käyttöön edellytetään lupa, mutta ilman lupaa arkistoa saavat käyttää nimetyt henkilöt. Arkiston käytölle on aikarajoitus: arkiston käyttö on kielletty xx.xx.xxxx saakka. Sen jälkeen arkisto on vapaasti käytettävissä. Arkiston käytölle on aikarajoitus: Lupa arkiston käyttöön on xx.xx.xxxx saakka. Sen jälkeen...[ehdot vaihtelevat] Arkistosta saa käyttää vapaasti jotain osaa, mutta osa arkistosta on suljettu joko pysyvästi tai määrätyn ajan. Arkiston käyttö rajoitetaan vain tieteelliseen tutkimukseen. Arkiston käyttö rajoitetaan vakavaan tieteelliseen tutkimukseen ja N.N.:n omaisille ja heille vain tiettyyn tarkoitukseen. Arkiston käytölle asetetaan jokin muu rajoitus.

5. Arkiston käyttö on rajoitettu arkistoteknisistä syistä. Ei koske sähköisiä aineistoja.

Yksityisarkistot

Suomessa toimii useita yksityisiä arkistoja jotka arkistoivat oman alansa yhteisöjen ja yksityishenkilöiden arkistoja ja antavat niitä tutkijoiden ja suuren yleisön käyttöön. Pysyväisluonteista valtionapua nauttavia yksityisiä arkistoja on yksitoista: Kansan Arkisto, Keskustan ja maaseudun arkisto, Porvarillisen työn arkisto, Suomalaisen Kirjallisuuden Seuran arkisto, Suomen Elinkeinoelämän keskusarkisto ELKA, Suomen Urheiluarkisto, Svenska Litteratursällskapet i Finlandin arkisto, Svenska Centralarkivet, Toimihenkilöarkisto, Työväen Arkisto ja Urho Kekkosen arkisto.

Arkistosektorilla toimii yllä mainittujen lisäksi muitakin organisaatioita, kuten Kansallisarkisto ja Musiikkiarkisto JAPA sekä kaupunginarkistot ja kotiseutuarkistot, jotka säilyttävät ja tarjoavat käyttöön yksityisiä arkistoaineistoja.

Luovutussopimukset ja käyttörajoitukset

Luovutussopimuksiin ja käyttörajoituksiin liittyvät käytännöt ovat olleet yksityisarkistoissa vaihtelevia. Luovutussopimuksia pyritään saattamaan yhtenäisempään muotoon esimerkiksi ottamalla käyttöön Creative Commons -lisenssejä. CC-lisenssien avulla ei kuitenkaan pystytä määrittelemään kuin osa arkistoaineistoon kohdistuvista käyttörajoituksista. Ja mikäli arkisto ei omista aineiston tekijänoikeuksia, se ei myöskään pysty niistä luopumaan CC-lisenssillä.

Aineistoluovutusten yhteydessä tehdään aineiston käytöstä ja oikeuksista sopimus luovuttajan kanssa. Sopimusta ei aina pystytä tekemään, mikäli luovuttajan suhde aineistoon on epäselvä: luovuttajalla on esimerkiksi niin sanottujen kolmansien osapuolien aineistoa hallussaan tai hänellä ei ole valtuuksia allekirjoittaa sopimusta. Tällaisia aineistoja ovat muun muassa haastatteluaineistot, joita tehtäessä luovuttajalla ei ole välttämättä ollut selvillä, että haastattelut tullaan myöhemmin luovuttamaan arkistoon.

Aineistoa saatetaan kuitenkin ottaa vastaan, vaikka luovutussopimusta ei voitaisikaan tehdä, mikäli aineiston kulttuurihistoriallinen merkitys on keskeinen. Myös esimerkiksi toimintansa lopettaneiden yhteisöjen aineistoa saatetaan ottaa vastaan ilman luovutussopimusta, sillä tällöin ei ole enää olemassa sopimuskumppania, jonka kanssa sopimus voitaisiin allekirjoittaa. Luovutussopimuksissa olisikin hyvä määritellä, miten toimitaan siinä tapauksessa, kun luovuttavaa organisaatiota ei enää ole olemassa.

Luovutussopimuksessa määritellään, onko aineisto vapaasti käytettävissä vai kohdistuuko siihen käyttörajoituksia, jotka voivat koskea koko aineistoa tai vain osaa siitä. Yleisesti pyritään siihen, että aineistot olisivat vapaasti käytettävissä. Toisinaan luovuttaja kuitenkin haluaa kieltää aineistojen kaupallisen käytön, jollei siitä sovita tapauskohtaisesti erikseen. On myös mahdollista, että arkisto suljetaan luovutussopimuksen perusteella määräajaksi. Joissakin tapauksissa luovuttaja haluaa kieltää kaiken käytön ilman, että siitä sovitaan erikseen. Tällaisia aineistoja ei pääsääntöisesti oteta vastaan.

Vanhoissa luovutussopimuksissa ei ole osattu varautua käyttöön tietoverkkojen välityksellä. Mikäli luovuttajataho on edelleen olemassa, niin vanhoja luovutussopimuksia voidaan saattaa ajan tasalle esimerkiksi samassa yhteydessä kuin kokoelma karttuu.

Osa yksityisarkistojen materiaalista saadaan aktiivisen hankinnan tuloksena kuten perinne- ja muistitietokeruiden sekä kenttätöiden ja haastatteluiden kautta. Näiden osalta kirjallisia sopimuksia ja suostumuksia on tehty enemmän määrin vasta 1990-luvun lopulta alkaen. Usein muistitietoaineistot on tallennettu ja luovutettu arkistolle vain tiettyyn käyttötarkoitukseen kuten tutkimuskäyttöön.

Jos aineistolla on käyttörajoitus, yhteyshenkilö (luvanantaja, usein sukulaiset tai jälkeläiset tai arkistonjohtaja) sekä käyttörajoituksen päättymisajankohta kirjataan sopimukseen. Osa kokoelmista saattaa olla luovuttajan pyynnöstä salaisia, jolloin niiden metatiedotkaan eivät ole julkisia. Salassapidon päättymisaika määritellään luovutussopimuksessa.

Yksityisarkistoissa pyritään siihen, että aineiston omistusoikeus siirtyy arkistolle. Joissakin arkistoissa lähdetään kuitenkin siitä, että edelleen toimivalla arkistonmuodostajalla on tietty omistusoikeus omaan ja edeltäjäorganisaatioiden arkistoon.

Jos aineistolla on useampia oikeudenomistajia, tulee vaatia valtakirja kaikilta oikeudenomistajilta. Mikäli luovuttajalla on tekijänoikeus aineistoon, hän voi antaa luovutussopimuksessa arkistolle oikeuden digitalisoida, muokata, kääntää ja julkaista aineistoa sekä tietoa aineistosta.

Julkisuuslain salausmääräykset koskevat myös yksityisiä toimijoita, jos niiden hallussa on esimerkiksi kopio viranomaisen salassa pidettäväksi määritellystä asiakirjasta. Myös esimerkiksi pankkisalaisuutta koskevien asiakirjojen tutkimuskäyttöä on rajoitettu.

Arkistonmuodostaja voi toisinaan luovuttaa aineistoa digitoitavaksi siten, että fyysinen aineisto palautetaan lahjoittajalle, kun taas digitoidun aineiston omistusoikeus siirtyy arkistolle.

Henkilötiedot ja käyttörajoitukset

Kaikki yksittäisiä ihmisiä koskevat tiedot ovat henkilötietoja, joiden keräämistä ja käsittelyä säätelee Henkilötietolaki. Arkaluonteisina tietoina pidetään henkilötietoja, jotka kuvaavat esimerkiksi rotua, etnistä alkuperää; yhteiskunnallista, poliittista, uskonnollista vakaumusta; rikollista tekoa, rangaistusta; terveydentilaa, sairautta, vammaa; seksuaalista suuntautumista; sosiaalihuollon tukitoimia tai etuuksia. Yksityisarkistojen aineistot sisältävät runsaasti henkilötietoja, jotka ovat olennaisia aineiston tieteellisen ja kulttuurihistoriallisen arvon kannalta. Nämä tiedot voidaan usein määritellä arkaluonteisiksi, koska ne liittyvät yksityisiin elämäntilanteisiin. Keruissa aineistojen aihe voi olla jo lähtökohtaisesti arkaluonteinen ja kuulua henkilötietolain 3 luvun 11 §:ssä yllä mainittuihin ryhmiin.

Henkilötietolaki asettaa arkaluonteisia henkilötietoja sisältävien aineistojen käytölle rajoituksia: niitä voidaan antaa käyttöön määrätarkoituksiin (tavallisimmin määritetty sanalla "tutkimuskäyttö"), kontrolloidusti (yleisesti arkiston tiloissa siten, että aineiston jäljentämistä on rajoitettu) ja sitoumusta vastaan (tunnistettu käyttäjä, allekirjoitus). Merkittävä osa aineistoista on siten käyttörajoitettuja tai näyttörajoitettuja (metatietojen näyttämiseen sisältyy rajoituksia). Luvan aineiston käyttöön myöntää yleensä arkistonjohtaja tai sopimuksessa mainittu taho kuten luovuttaja.

Metatiedot ja näyttörajoitukset

Metatiedot sisältävät henkilötietoja, jolloin niistä muodostuu henkilörekistereitä, joiden näyttäminen tietoverkoissa on ongelmallista. Esimerkiksi suku, perhe- ja henkilöarkistojen metatietoihin saattaa sisältyä näyttörajoitus.

Osalla yksityisarkistoista on tällä hetkellä arkistoluettelot - myös henkilöarkistojen osalta – luettavissa internetissä. Uudisteilla olevan arkistolainsäädännön toivotaan selkeyttävän yksityisarkistojen luetteloiden julkisuusrajoituksia.

Myös tiedot aineistojen luovuttajista ja keruiden vastaajista, perinteen tallentajista, haastateltavista ja valokuvien kuvatuista henkilöistä voivat olla näyttörajoitettuja. Lisäksi jotkut henkilöistä ovat erikseen asettaneet rajoituksen nimensä saattamisesta julkisuuteen tai ovat halunneet näiden tietojen päätyvän vain tutkimuskäyttöön. Metatietojen näyttörajoitukset voivat liittyä myös arkistovirkailijan tekemiin ratkaisuihin tai luovuttajan asettamiin rajoituksiin.

Osa luovutuksista on tapahtunut vuosikymmeniä sitten, jolloin ei ole vielä keskusteltu henkilörekisteristä. Nykyisetkään luovuttajat välttämättä ajattelevat nimistään muodostuvan henkilötietolainsäädännön mukaista henkilörekisteriä, vaikka he haluaisivat nimensä olevan arkistossa ja löytyvän arkiston hakujärjestelmästä.

Näyttörajoitettuja metatietoja voi käyttää vain asiakastiloissa käyttösitoumusta vastaan. Käyttösitoumuksessa käyttäjä sitoutuu muun muassa noudattamaan tietoja antaneiden henkilöiden asettamia rajoituksia.

Tekijänoikeus ja käyttörajoitukset

Yksityisarkistojen aineistot sisältävät runsaasti tekijänoikeuden alaista aineistoa. Tällaisen aineiston näyttäminen internetissä on kielletty ilman oikeudenomistajan erityistä lupaa. Tekijänoikeuden piirissä olevia digitaalisia aineistoja voidaan kuitenkin näyttää arkiston tiloissa tekijänoikeuslain määrittämin määräedellytyksin. Mikäli aineistoja halutaan käyttää arkiston ulkopuolella esimerkiksi esitys- tai julkaisutarkoituksessa, tulee asiakkaalla olla luvat kaikilta oikeudenhaltijoilta ennen kuin hän saa aineiston käyttöönsä. Yksityisarkistoilla saattaa myös olla museoiden tapaan maksullisia kuvamyyntipalveluita niiden kuvien osalta, joiden oikeudet arkisto omistaa tai jotka ovat tekijänoikeudellisesti vapaita tai jos on erillinen sopimus tekijänoikeudenomistajan kanssa.

Osassa yksityisarkistoja tekijänoikeudet ovat merkittävä este aineistojen laajalle avaamiselle tietoverkkoon. Esimerkiksi Musiikkiarkisto JAPAN aineistoista noin 85 % on tekijänoikeuden alaista materiaalia, ja lähes poikkeuksetta JAPAlle ei ole luovutettu tekijänoikeutta aineistoon.

Mikäli tekijä on siirtänyt teoksensa taloudelliset oikeudet jollekin toiselle osapuolelle, täytyy arkiston tehdä tekijänoikeudelliset sopimukset kyseisen toimijan kanssa. Tekijänoikeuksien luovuttamista rajoittavat myös tekijän sopimukset tekijänoikeusjärjestöjen kanssa. Tekijänoikeuksien selvittäminen on usein työlästä varsinkin kuolleiden tai tuntemattomien tekijöiden osalta.

Esimerkiksi valokuvien käyttöä arkistoissa hankaloittaa se, että teoskynnys täytyy määritellä jokaisen kuvan osalta erikseen. Mikäli kuva ylittää teoskynnyksen, sen suoja-aika on merkittävästi pitempi. Toinen tekijänoikeuksiin ja luovutussopimukseen liittyvä ongelma on se, että luovuttajalla on usein tekijänoikeus aineistoon vain osittain (esimerkiksi musiikkiyhtyeen johtaja saattaa luovuttaa yhtyeensä äänitteitä arkistoon, mutta hän ei omista muiden muusikoiden tekijänoikeuksia).

Kirjastot

Kokoelmat

Kirjastojen kokoelmiin hankitut painetut aineistot ovat perinteisesti olleet kaikkien asiakkaiden lainattavissa. Käyttöoikeuksia (koskien esimerkiksi lainauskäytäntöjä, valokopiointia, ja skannausta) ei ole ollut tarpeen tallentaa yksittäisten julkaisujen metatietoihin, vaan asiakkaita on informoitu niistä yleisesti. Mahdolliset käytön rajoitukset ovat yleensä liittyneet erikoiskokoelmiin (esim. käsikirjastoaineistoa ei lainata) tai julkaisujen huonoon kuntoon, joskus myös aineiston ei-julkaisuuteen tai salaisuuteen.

Elektronisia aineistoja (CD ROM -levyt, etäkäyttöiset tietokannat) alettiin hankkia kirjastoihin jo 1980-luvulla. Elektronisten kokoelmien merkittävä kartunta alkoi kuitenkin vasta 1990-luvun puolivälin jälkeen ensimmäisten kaupallisten lisenssisopimusten myötä. Samaan aikaan käynnistyivät myös digitointihankkeet, joissa kirjastot saattoivat käyttäjien ulottuville etupäässä tekijänoikeuksista vapaata aineistoa. Esimerkiksi tekijänoikeudesta vapaat Kansalliskirjaston digitoimat aineistot ovat verkossa vapaasti käytettävissä. Niitä voi lukea, kuunnella, tulostaa tai ladata itselleen kuka tahansa, mistä tahansa ja milloin tahansa. Digitaalisten aineistojen osuus kaikista hankituista aineistoista on nopeasti kasvanut, ja on arvioitu että vuonna 2020 ne muodostavat tieteellisissä kirjastoissa noin 80 % kaikista hankituista aineistoista.

Kaupallisten kustantajien julkaisemia tieteellisiä kausijulkaisuja ja nyttemmin myös monografia-aineistoja koskevat lisenssit eivät kuitenkaan ole edellyttäneet kirjastoilta (tai kustantajilta) ja niiden käyttämiltä järjestelmiltä käyttöoikeuksien metatietojen tallentamista ja hyväksikäyttöä. Kustantajien suosimat lisenssiratkaisut ovat olleet käyttöoikeuksien kannalta yksinkertaisia, mikä on ollut sekä kustantajille, kirjastoille että käyttäjille edullinen ratkaisu. Kustantajat eivät myöskään ole halunneet profiloitua käyttöoikeuksien valvojina enempää kuin on pakko. Käytännössä nykytilanteessa siis vain rajoitetaan pääsyä lisensoituihin aineistoihin organisaatiokohtaisten käyttäjätunnuksien avulla. Mutta nykytilanteessa käyttäjien voi olla vaikea selvittää esimerkiksi johonkin lisensoituun tieteelliseen artikkeliin liittyviä käyttöoikeuksiaan tai oikeuksien omistajia, olipa artikkeli poimittu tieteellisestä lehdestä tai tekijän taustaorganisaation julkaisuarkistosta. Järjestelmissä ei ole käyttöoikeuksien selvittämiseen tarvittavia metatietoja, eikä organisaatioiden taustajärjestelmissä tähän tarvittavia teknisiä ominaisuuksia.

Julkaisujen käyttöoikeuksien kuvaaminen voi olla vaikeaa. Vaikka teoksen tekijänoikeudellinen status on aina sama, sen eri kappaleilla voi olla eri käyttöoikeudet, ja ohjelmistojen tulee ottaa tämä huomioon. Elektronisesta väitöskirjasta voi olla olemassa kansalliskirjastoon luovutettu ja vain hyvin rajatusti käytettävissä oleva elektroninen vapaakappale, väittelijän yliopiston julkaisuarkistossa oleva vapaasti käytettävissä oleva kappale sekä kustantajan palvelimella oleva maksullinen kappale, joka on voitu lisensoida esimerkiksi yliopistojen henkilökunnan ja tutkijoiden käyttöön, mutta ei samoin oikeuksin kuin julkaisuarkistossa oleva kappale.

Nykyistä teknistä takapajuisuutta selittää osaltaan se, että kirjastojen soveltamaan MARC 21 -formaattiin lisättiin käyttöoikeuksien metatietojen edellyttämiä kenttiä vasta 2000-luvulla. Niiden käyttö on toistaiseksi ollut varsin vähäistä paitsi Suomessa myös muualla. Tavoitteena kuitenkin on, että jatkossa asiakkaille kerrotaan esimerkiksi teoksen tekijänoikeudellinen status sekä - jos aineisto on tekijänoikeuden suojaama - mahdollinen käytön salliva lisenssi ja sen sanelemat ehdot. Samat tiedot pitäisi kertoa myös sovelluksille koneymmärrettävässä muodossa. Tärkeää on ilmaista myös esim. public domain -fraasilla että (digitoitu) aineisto on täysin vapaasti jatkokäytettävissä.

Kuvailevan metatiedon asiantuntijaryhmä KUMEA ylläpitää MARC 21 -formaattiin perustuvaa kansallista käyttöoikeuksien metatietojen tallennusohjetta. Ohje täydentää MARC-formaattia ja pyrkii yhdenmukaistamaan kirjastojen kuvailukäytänteitä. Se on kaikkien kirjastojen ja muidenkin kiinnostuneiden käytettävissä verkossa.⁶ Ohjeeseen ollaan tätä kirjoitettaessa lisäämässä tutkimusdatan käyttöoikeuksien kuvailun edellyttämiä lisäominaisuuksia.

Liitteessä kuvataan tarkemmin ohjeen mukaisesti tallennettavien käyttöoikeuksien metatietojen rakennetta.

Julkaisujen metatiedot ja henkilötiedot

Kirjastojen kansallinen ja kansainvälinen kuvailuyhteistyö perustuu julkaisujen metatietojen kansainväliseen vaihtoon. Suomessa uutuushankinnoista kuvaillaan jopa yli 90 % poimimalla kuvailutiedot vapaasti käytettävissä olevista muiden kirjastojen rekistereistä standardirajapintojen kautta. Suomalaisista tietokannoista muualle poimittavista tietueista ei poisteta eikä yleensä edes karsita henkilötietoja (julkaisujen tekijät, kohdehenkilöt), joiden tarkkuuden on oltava riittävä yksiselitteiseen identifiointiin. Tämä on erittäin tärkeää esimerkiksi tieteellisen meriitin ja tekijänoikeuksien kohdistamiseksi.

⁶ Käyttöoikeuksien metatieto Kuvailevan metatiedon asiantuntijaryhmän wikissä, <https://www.kiwi.fi/x/i4VDAq>

Toimijoita koskevat tiedot voidaan koota erillisiin auktoriteettitietokantoihin. Kansalliskirjasto ylläpitää kirjastojen yhteistä Asteri-tietokantaa, jossa on yli 120 000 henkilön ja yli 40 000 organisaation tietoja (minimissään pelkkä nimi).

Museot

Yleistä museoiden tietovarantojen käyttöoikeuksista

Museoille on kertynyt 1800-luvulta lukien valtava määrä erilaisia tietovarantoja, jotka ovat syntyneet ennen nykyisen lainsäädännön asettamia vaateita mm. henkilötietojen tallentamisen suhteen. Erilaisia vastaanotto- ja niihin liittyviä sopimuskäytäntöjä on ja on ollut paljon. Tähän vaikuttaa mm. museosektorin monimuotoisuus ja eri aikoina vallalla olleet yleiset käytännöt. Käyttöoikeuksiin liittyvä problematiikka on nykyään verkkosaatavuuden myötä täysin toinen kuin fyysisten tietovarantojen aikakaudella. Aikaisemmin museon kokoelma-amanuenssille riitti, että käyttöoikeudet voitiin määritellä yksittäisiin tietopalvelupyyntöihin tapauskohtaisesti nojaamalla museon sisäisiin käytäntöihin. Nykyään aineiston käyttöoikeudet ja -rajoitukset täytyisi pystyä määrittelemään yksiselitteisesti kenelle tahansa missä tahansa ja mielellään kansainvälisesti ymmärrettävässä muodossa.

Museosektorilla on tällä hetkellä käytössä kymmeniä kokoelmahallintajärjestelmiä, joiden suhteen ei ole keskitetysti selvitetty miten niiden tietomalleissa on huomioitu aineistojen käyttöoikeuksiin liittyvien tietojen tallentaminen.⁷ Museoiden tilanne poikkeaa myös esimerkiksi kirjastoista siinä, ettei tietojen tallentamiseen ole käytössä yhteistä formaattia (vrt. MARC 21 kirjastoissa). Museoiden suhteen on myös huomioitava, ettei myöskään luettelointia ole ohjeistettu kattavalla ja yhtenäisellä luettelointiohjeella vasta kuin parin viime vuoden aikana. Tästä syystä museoiden näkökulmasta tärkeää olisi ennen kaikkea sekä kokonaiskuvan tarkempi selvittäminen että riittävän yleisen tason linjausten aikaansaaminen ja toimivien yhteisten käytäntöjen jalkauttaminen yksittäisten museoiden toimintaan.

Viime vuosina museosektorin tilanne on kuitenkin muuttunut monessa suhteessa positiiviseen suuntaan: valtakunnallisessa Museo 2015-yhteishankkeessa on kehitetty museoiden sähköistä kokoelmahallintaa mm. luetteloinnin, tiedonhallinnan prosessien ja välineiden (kokonaisarkkitehtuuryö) sekä verkkosaatavuuden näkökulmasta (Museoiden Finna).⁸ Museo 2015-hankkeessa työstetty kansainväliseen Spectrum 4.0-standardiin perustuva Museoiden luettelointiohje⁹ tarjoaa esineen, valokuvan, taideteoksen, arkistoaineiston ja audiovisuaalisen aineiston oikeuksiin (mm. tekijänoikeudet) ja objektin käyttämiseen liittyvien tietojen (mm. käyttörajoitukset) tallennusohjeet. Lisäksi luettelointiohjeessa on huomioitu objektiin liittyvät vaatimukset kuten turvallisuus- ja käsittelyohjeet. Museoiden luettelointiohje kertoo miten nämä tiedot tulee tallentaa, mutta ei sinänsä ohjeista millä tavalla esimerkiksi lisensoijia tulisi soveltaa museoiden tuottamiin aineistoihin.

Käyttöoikeudet hankinnan näkökulmasta

Museoiden kokoelmat karttuvat kokoelmakehittämisen ydinprosessissa eli hankinnassa.¹⁰ Hankinta pitää sisällään eri hankintatavat kuten mm. lahjoituksen, oston ja deponoinnin. Museoiden pyrkimyksenä on aina selvittää hankinnan yhteydessä sitä koskevat oikeudet. Hankinnan yhteydessä luovuttajan kanssa sovitaan kirjallisesti siihen liittyvien oikeuksien luovutuksesta. Kaikissa

⁷ Ks. museoiden kokoelmahallintajärjestelmistä tarkemmin Museoiden kokoelmahallinnan kokonaisarkkitehtuuri 1.0. Museoviraston oppaita ja ohjeita 2. Helsinki 2013.

⁸ Museo 2015 http://www.nba.fi/fi/museoalan_kehittaminen/museo_2015

⁹ Luettelointiohje www.luettelointiohje.fi

¹⁰ Hankinnan toisena ydinprosessina kokoelmakehittämisen yläprosessina voidaan pitää poistoprosessia. Ks. poistoprosesseista tarkemmin Kokoelmapoistojen hyvät käytännöt. Sarantola-Weiss, Minna & Västi, Emilia (toim.). Suomen museoliitto 2015.

http://www.museoliitto.fi/doc/verkkajulkaisut/Kokoelmapoistojen_hyvät_kaytannot.pdf

tapauksissa tämä ei kuitenkaan ole mahdollista, jos luovuttajalla ei ole kuin omistusoikeus luovutettavaan aineistoon. Tässä suhteessa museot ovat usein hankintojen suhteen samassa tilanteessa kuin esimerkiksi yksityisarkistot. Eräänä hankintatapana on huomioitava myös museoiden toteuttamat nykydokumentointihankkeet (keruut), joiden yhteydessä tulee selvittää kerättyyn aineistoon liittyvät oikeudet ennen kaikkea henkilötietolain (koskien erityisesti arkaluonteisia tietoja) ja tekijänoikeuslainsäädännön näkökulmasta.

Kokoelmia vastaanotettaessa laaditaan nykyään yleisesti luovutus sopimus tai lahjakirja museossa käytössä olevalle pohjalle, jossa määritellään myös aineistojen käyttöä. Sopimusmallit eivät ole määrämuotoisia eivätkä standardoituja, joten monissa museoissa sopimuksissa ei ainakaan vielä määritellä oikeuksia ja mahdollisia rajoituksia aineiston jatkokäytön kannalta kovin laajasti. Tästä syystä tässä selvityksessä ei voida tarkastella miten yksittäisissä museoissa huomioidaan aineistoluovutuksen yhteydessä laadituissa sopimuksissa sitä koskevat oikeudet. Yleisesti ottaen kokoelmien käyttöön liittyvät käytännöt noudattavat arkistolaitoksen käytäntöjä, mutta eivät ole vielä yhtä vakiintuneita. Tulevaisuuden kannalta museoiden näkökulmasta olisi tästä syystä tärkeää selkeyttää ja harmonisoida sopimuskäytäntöjä osana aineistojen käyttöoikeuksien kuvailun yhtenäistämistä.

Käyttöoikeuksiin liittyen on huomioitava myös ennen nykyistä verkkosaatavuutta tehdyt hankintasopimukset. Vaikka luovuttaja on luovuttanut aineiston "täysin oikeuksin" museoon, hankintahetkellä hän on voinut olla siinä luulossa, että aineisto tulee olemaan pääosin museon sisäisessä käytössä ja asiakkaiden saatavilla ainoastaan kontrolloidusti museon asiakaspalvelu- tai näyttelytiloissa. Myös erilaisiin suullisiin sopimuksiin tai vain vakiintuneisiin käytänteisiin ("näin on näiden aineistojen suhteen aina toimittu") perustuvat hankintasopimukset tulisi pyrkiä jälkikäteen saattamaan kirjalliseen muotoon. Tämä on monissa tapauksissa tosin mahdotonta pitkien aikojen jälkeen, joten yleisesti museoissa noudatetaan tulkintaa, että tekijänoikeuksista vapaa aineisto on käytettävissä museon itsensä määrittelemillä ehdoilla.

Henkilötietolaki museoiden näkökulmasta

Museoiden aineistot sisältävät runsaasti henkilötietoja, jotka ovat olennaisia niiden tutkimuksellisen ja kulttuurihistoriallisen arvon ja merkittävyyden kannalta. Tiedot sisältävät myös arkaluonteisiksi määriteltyjä asioita. Arkaluonteisuus koskee erityisesti kuva-aineistoja, mutta myös muita museoiden aineistoja. Lisäksi eri aikoina käytössä olleet luettelointitavat ja arvostukset näkyvät edelleen metatiedoissa. Metatietojen historiallinen kerroksisuus on pyrittävä säilyttämään museoissa myös jatkossa, vaikka nykynäkökulmasta metadata on eri syistä ongelmallista tai puutteellista. Esimerkiksi museoetiikan näkökulmasta osa näistä vanhoista tiedoista voi olla nykyään eri syistä loukkaavaksi katsottavaa.

Edellä mainituista syistä erityisenä haasteena museoiden säilyttämien aineistojen suhteen on henkilötietolaki (22.4.1999/523). Henkilötietolaissa todetaan (3§), että "henkilötiedolla kaikenlaisia luonnollista henkilöä taikka hänen ominaisuuksiaan tai elinolosuhteitaan kuvaavia merkintöjä, jotka voidaan tunnistaa häntä tai hänen perhettään tai hänen kanssaan yhteisessä taloudessa eläviä koskeviksi." Henkilörekisterillä puolestaan tarkoitetaan käyttötarkoituksensa vuoksi yhteenkuuluvista merkinnöistä muodostuvaa henkilötietoja sisältävää tietojoukkoa, jota käsitellään osin tai kokonaan automaattisen tietojenkäsittelyn avulla taikka joka on järjestetty kortistoksi, luetteloksi tai muulla näihin verrattavalla tavalla siten, että tiettyä henkilöä koskevat tiedot voidaan löytää helposti ja ilman kohtuuttomia kustannuksia. Esimerkiksi valokuva-aineistojen suhteen on tulkittu, että jos henkilö on tunnistettavissa valokuvasta, katsotaan sen olevan laissa tarkoitettu henkilötieto.

¹¹ Tämä on huomioitava erityisesti luetteloinnissa ja mahdollisten

¹¹ Valokuva ja yksityisyyden suoja henkilötietolain kannalta. Päivitetty 27.7.2010
Tietosuojavaltuutetun toimisto.

käyttörajoitusten merkitsemisessä. Samanaikaisesti tällä on myös vaikutusta museoiden aineistojen verkkosaatavuuteen: henkilötietolaki määrittää mitä aineistoja museo ylipäätään voi asettaa verkkoon asiakkaiden selailtavaksi.

Henkilötietolaissa museot eivät ole millään tavalla erityisasemassa esimerkiksi kaupallisiin toimijoihin nähden. Henkilötietolaissa todetaan (3§), että "rekisterinpitäjällä tarkoitetaan yhtä tai useampaa henkilöä, yhteisöä, laitosta tai säätiötä, jonka käyttöä varten henkilökisteri perustetaan ja jolla on oikeus määrätä henkilökisterin käytöstä tai jonka tehtäväksi rekisterinpito on lailla säädetty". Tällaista erityistä tehtävää ei ole lainsäädännössä määritelty eksplisiittisesti museoille. Tosin Museolaissa (3.8.1992/729, 1§) todetaan muun muassa, että "museoiden tulee edistää kulttuuri- ja luonnonperintöä koskevan tiedon saatavuutta tallentamalla ja säilyttämällä aineellista ja visuaalista kulttuuriperintöä tuleville sukupolville." Tiukasti tulkiten henkilötietolaki johtaa siihen, että vähänkään uudemman ajan aineistosta vain maisemakuvia ja muita ilman mitenkään tunnistettavissa olevia ihmisiä olevia kuvia on mahdollista avata verkkoon saavutettavaksi. Tämä rajoittaisi erittäin merkittävästi sekä museoille määriteltyjä tehtäviä että laajemmin myös tutkimusta. Tästä syystä on tärkeää, että aloitetaan henkilötietolain uudistaminen sillä tavoin, että se mahdollistaa kulttuuriperintö- ja tutkimusaineistojen laajemman käytön. Tätä tavoitetta voidaan edistää mm. KAM-juridiikkaryhmän vetämänä.

Museoiden aineistojen käyttörajoituksista

Museoiden luettelointiohje ohjeistaa luetteloimaan käyttörajoitukset yksittäisellä termillä. Käytetyt termit ovat museokohtaisia eivätkä koko museokenttää koskevia. Käyttörajoituksen yhteyteen on merkittävä aina tieto siitä milloin käyttörajoitus on kirjattu. Järjestelmään voidaan jättää myös aiemmat jo rauenneet käyttörajoitukset, jos niiden kirjaamisaika on merkitty, sillä tällä tavoin ne voidaan erottaa toisistaan.

Kokoelmat ovat lähtökohtaisesti julkisia, mutta niihin voi sisältyä sekä pääsy- että käyttörajoituksia. Erilaiset pääsy- tai käyttörajoitukset voivat perustua suoraan luovuttajan määrittämiin ehtoihin tai museon henkilökunnan lainsäädännön ja eettisten näkökohtien perusteella asettamiin ehtoihin. Esimerkiksi Museovirastossa on käytössä käyttöluvahakemus, jolla pyydetään tutkittavaksi arkeologisia esinekokoelmia sekä tietosuojasysteistä tai aineiston huonokuntoisuuden vuoksi käytöltään rajoitettuja kuva- tai arkistoaineistoja ja julkaisuja. Käyttöluvahakemukset hyväksyy kustakin aineistoista vastaavan yksikön päällikkö.

Museot eivät ole asettaneet verkkoon saataville sellaista metadataa, johon pääsyä olisi syytä rajoittaa esimerkiksi tunnistautumisen avulla. Museoiden verkkoon saataville asetettavan metadatan suhteen on huomioitava, että LIDO-mappauksessa museot voivat määrittää mitkä kokoelmahallintajärjestelmän kentät mapataan. Luetteloihin tulee museoissa ohjeistaa erityisen tarkkuuteen näiden kenttien suhteen, jotta julkisesti saataville ei päädy esimerkiksi arkaluonteiseksi määriteltäviä tietoja. Suurin osa museoiden metadatasta on saatavilla vain kokoelmahallintajärjestelmässä eikä sitä aseteta verkon kautta saataville.

Monet museot tarjoavat myös käyttörajoitusten mukaan pääsyn suoraan museon kokoelmanhallintajärjestelmään asiakaspalvelutiloissa. Museot eivät ole useimmiten vieneet esimerkiksi tekijänoikeuslain asettamien käyttörajoitusten alaisista aineistoista metadataa verkkoon, jos ilmentymää ei ole voitu viedä em. systä selailtavaksi. Tällaisissa tapauksissa asiakkaat pääsevät tutustumaan museon asiakaspalvelutiloissa sellaiseen aineistoon, jota ei ole viety verkkoon selailtavaksi. Kuitenkin esimerkiksi henkilötietolain (22.4.1999/523) 13§:n määrittämiin arkaluonteisiin tietoihin ei ole asiakkailta pääsyä kuin

http://www.tietosuoja.fi/material/attachments/tietosuojavaalutettu/tietosuojavaalutetuntoimisto/opaat/6JfqOQLTj/Valokuva_ ja_ yksityisyyden_ suoja_ henkilotietolain_ kannalta.pdf

käyttötarkoituksen (esim. tutkimuskäyttö) perusteella ja kirjallista sitoumusta (käyttölupa) vastaan. Toiseksi erityisesti taidemuseokentällä on huomioitava teosturvallisuuteen liittyvät vaateet: ulkopuolisten tietoon ei tule välittää esimerkiksi yksityisessä omistuksessa olevan teoksen omistussuhteita koskevaa metadataa.

Ilmentymien käyttöoikeuksista

Finnan myötä tavaksi on tullut erottaa objektin ilmentymien käyttöoikeudet museosektorilla seuraavalla tavalla: Finnan kautta saatavilla olevat esikatselu- ja näyttökuvat lisensoidaan nykyään useimmiten Creative Commons -lisenssejä hyödyntäen, mutta painokelpoisia digitaalisia kuvia koskevat käyttöehdot ovat edelleen organisaatiokohtaisesti määriteltyjä ja siten hyvinkin kirjavia. Jälkimmäinen käytäntö liittyy museoiden kuvamyyntipalveluihin. Toistaiseksi täysresoluutiokuvia ei ole juuri muutamia Flickr-avauksia lukuun ottamatta saatettu verkkoon Creative Commons-lisensseillä. Asiakkaat voivat sen sijaan hankkia painokelpoisia digitaalisia kuvia maksua vastaan eri käyttötarkoituksiin joko suoraan museosta tai verkkoon toteutetusta kuvamyyntipalvelusta.¹² Näiden aineistojen hinnoittelu- ja käyttöoikeuskäytännöt ovat museokohtaisia.

Museoalalla on käyty viime aikoina runsaasti keskustelua museoiden kuvamyyntipalveluiden tulevaisuudesta, mutta selkeää yhtenäistä koko museokentän läpäisevää näkemystä tulevaisuuden vaihtoehtoista ei ole. Toisaalta kuvamyyntiin kohdistuu entistä voimakkaampia muutospaineita aineistojen avaamisen näkökulmasta, mutta yhtäältä museoiden taloudelliset resurssit asettavat reunaehdot maksullisten aineistojen avaamiselle.

Tiivistetysti erilaisten tietovarantojen avaamisessa on kysymys laaja-alaisesta sekä julkishallinnon että kansalaisten toimintakulttuurin muutoksesta. Toimintaympäristön nopeat muutokset tulee huomioida myös museoiden ylläpitämien digitaalisten aineistojen käyttöoikeuksia määriteltäessä. PSI (Public Sector Information) -direktiivin laajennus edellyttää, että museoiden, arkistojen ja kirjastojen aineistojen tulee olla vapaasti uudelleen käytettävissä, jos mikään lainsäädännöllinen näkökohta ei sitä eksplisiittisesti estä. Kuitenkin on huomioitava, että kulttuuriperintöaineistoja säilyttäville valtion laitoksille (ja niihin rinnastettaville organisaatioille sekä myös yleisille kirjastoille) lainsäädännössä asetetut, aineistojen saatavuuteen liittyvät tavoitteet, ylittävät nykytilanteessa PSI-direktiivin minimitaso vaatimukset. Tällä hetkellä tarvetta säännösmuutoksille ei ole. Myös valtion maksuperustelaki (150/1992) on PSI-direktiivin mukainen.

Kansallinen digitaalinen kirjasto ja Finna

Kansallisen digitaalisen kirjaston asiakasliittymän Finna ei edellytä sopimustasolla palveluun osallistuvilta organisaatioilta, arkistoilta, kirjastoilta ja museoilta, että palvelun indeksiin haravoitava metadata sisältäisi oikeuksien hallintaan liittyvää metatietoa. Esimerkiksi Europeana, joka painottuu konseptiltaan Finnaa selkeämmin kulttuuriperintöaineistoihin, edellyttää, että palvelussa saatavilla olevien digitaalisten objektien käyttöoikeus on ilmoitettava tietyin, määritellyin Right Statement -vaihtoehtoin.¹³

Metadataan liittyvät oikeudet ja rajoitukset

Europeanan tavoin myös Finnan palvelukonseptiin kuuluu metadatan lisensointi CC0-lisenssillä. Asia on kirjattu Finnan palvelusopimukseen, jossa todetaan, että aineistoja Finnaan tuova organisaatio: "Vastaa siitä, että sillä on oikeus asettaa vapaasti kaikkien käytettäväksi ja tämän sopimuksen liitteen 1 mukaisesti

¹² Jälkimmäisestä esimerkkinä Museoviraston kuvakokoelmat.fi

¹³ Europeana: Available Right Statements

<http://pro.europeana.eu/share-your-data/rights-statement-guidelines/available-rights-statements>

Nämä Right Statements -vaihtoehdot ovat muuttuneet marraskuussa 2016 Europeana ja Digital Public Library of American (DPLA) RightsStatements.org -yhteistyön myötä. Ks.

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

lisensoitavaksi Finnaan toimittamansa metatieto.” Liite 1 määrittää metatiedon lisenssiksi CC0:n. Lisensointi CC0:lla koskee ensi sijassa Finnan paikallisen indeksin ja ns. kansallisen näkymän eli Finna.fi:n metadattaa, ei esim. yksittäisen kirjaston omaa Finna-näkymää, jossa hakujen kohteena saattaa olla myös kaupalliselta toimijalta hankitun ulkoisen indeksin (Primo Central Index) metatieto.

Finnan palvelusopimukseen on kirjattu myös: ”Mikäli Finnaan toimitettavaan metatietoon kuitenkin liittyy käyttörajoituksia, konsortion jäsenen tulee yksilöidä metatieto Kansalliskirjaston ohjeistamalla tavalla ja ilmoittaa asiasta Kansalliskirjaston antamien ohjeiden mukaan.” Kirjaus on tarpeen koskien ainakin korkeakoulukirjastojen kirjastojärjestelmistä Finnan paikalliseen indeksiin haravoitavia kansainvälisten e-kirja-aineistojen metatietoja, joihin saattaa sopimussyistä liittyä käyttörajoituksia. Kansalliskirjasto on laatinut ohjeen tällaisen käyttörajoitetun metatiedon yksilöinnistä.¹⁴ Yksilöinnin perusteella käyttörajoitettu metatieto rajataan pois Finna.fi-näkymästä ja siten myös näkymän avoimesta rajapinnasta (api.finna.fi). Ohje suosittaa käyttörajoitetun metatiedon kuvailua tietuekohtaisesti Kuvailevan metatiedon asiantuntijaryhmän ohjeistuksen mukaan, mutta Finna tukee myös kokoelmakohtaista käyttörajoitetun metatiedon yksilöintiä.

Jos osa objektia tai aineistoa kuvailevasta tiedosta on käyttörajoitettua, Finnaan osallistuvat museot ja arkistot ovat tyypillisesti huolehtineet, ettei tällaista metatietoa päädy Finnan indeksiin LIDO- tai EAD-mappauksen myötä. On kuitenkin esitetty, että tällaiseen metatietoonkin tulisi tarjota Finnasta pääsy esimerkiksi vahvan tunnistautumisen avulla. Joissain tapauksissa myös kirjastojen metatiedon kohdalla joidenkin kenttien päätyminen Finnan estetään jo ennen metatiedon haravointivaihetta.

Kirjastoista Finnaan haravoitavassa metatiedossa tiivistelmiä sisältävissä kentissä (MARC-formaatissa 520, Dublin Coressa dc.abstract) saattaa olla kolmannen osapuolen tuottamaa kertovaa tekstiä, jossa teoskynnys ylittyy ja johon kirjastolla ei ole tekijänoikeutta. Finnan käyttöehdoissa todetaan, ettei tällainen kuvailutieto ole vapaasti käytettävää, ja tiivistelmäkentät rajataan pois avoimen rajapinnan kautta CC0-lisenssillä jaeltavasta metadattasta. Samoin rajataan pois Finnan käyttäjien tuottamat, kertovaa tekstiä sisältävät kuvailutiedot.

Aineistoihin liittyvät oikeudet ja rajoitukset

Vaikka Finnassa ei edellytetä, että palvelun kautta saataville tuotaviin aineistoihin liittyviä oikeuksia tai rajoituksia tulisi kuvailla, muodostaa tällaisen kuvailutiedon puute Finnan kannalta käytettävyysongelman ja esteen aineistojen käytölle. Erityisen akuutti ongelma on verkossa saatavilla olevan aineiston kohdalla.

Käyttäjäpalautteessa on kiinnitetty huomiota varsinkin suoraan Finnan käyttöliittymässä saatavilla olevien visuaalisten verkkoaineistojen käyttöoikeustietojen puutteisiin. Tällaiset verkkoaineistot ovat tyypillisesti museoiden ja arkistojen digitoitujen aineistojen ns. esikatselukuvia, joihin ei enää kohdistu tekijänoikeuden suoja tai joiden oikeudet kuuluvat kuvat Finnaan tarjoavalle organisaatiolle tai jotka voidaan sopimuksen perusteella esittää Finnassa. Asia on noussut esiin myös Finnan käytettävyytsteissä: jos käyttöoikeustiedot puuttuvat tai ovat epäselvät, käyttäjät eivät tiedä, mitä aineistolla voi tehdä ja ohittavat aineiston. Digitoitujen aineistojen esikatselukuvien käyttöoikeuksiin liittyvä metatieto on kuitenkin viime aikoina yleistynyt huomattavasti. Erityisesti Finnaan osallistuvat museot ovat alkaneet sisällyttää tällaista käyttöoikeustietoa Finnaan haravoitavaan metadataan. Usein käyttöoikeudesta ilmoitetaan CC-lisenssein.

Finna.fi:ssä otettiin alkuvuodesta 2016 käyttöön ensimmäinen käyttöoikeustietoihin perustuva hakuominaisuus. Sallitut käyttötavat -hakurajain

¹⁴ Käyttörajoitetun metatiedon yksilöinnin ohje Finna-wikissä, <https://www.kiwi.fi/x/Kw0XAw>

kohdistuu LIDO- ja EAD-formaateista saataviin museoiden ja arkistojen digitaalisten aineistojen käyttöoikeustietoihin eli käytännössä esikatselukuvien käyttöoikeuksiin.¹⁵ Jatkossa on tarkoitus ottaa myös kirjastojen vapaasti verkossa saatavilla olevien digitaalisten aineistojen käyttöoikeusmetadatan rajaimen avulla toteutettavien hakujen piiriin. Tällaisista aineistoista ei usein ole saatavilla esikatselukuvia Finnan käyttöliittymässä, joten rajain tulee jatkossa koskemaan myös sellaista digitaalista aineistoa, jonka tarkastelua varten siirrytään Finnasta linkin kautta toiseen käyttöliittymään. Laajennuksen jälkeen rajaimella voi hakea myös mm. teksti- ja ääniaineistoja.

Verkossa saatavilla olevien aineistoihin liittyvän pääsyräjoituksen ilmoittaminen metadatatassa olisi Finnan kannalta hyödyllistä. Käytettävyyden kannalta olisi tarpeellista, että pystyttäisiin erottelmaan vapaasti saatavilla oleva digitaalinen aineisto kirjautumista tai muuta tunnistautumista edellyttävästä aineistosta. Jatkossa Finnassa pyritään huomioimaan myös muita pääsyyn, saatavuuteen ja käyttöön liittyviä metatietoja, kuten fyysisiin aineistoihin liittyviä oikeuksia ja rajoituksia, ja vähintään esittämällä tällaiset metatiedot tarkoituksenmukaisesti käyttöliittymässä.

Digitaalisten aineistojen käyttöoikeustietojen merkitys on korostunut Finnassa entisestään avoimen rajapinnan käyttöönoton myötä. Rajapinnan kautta on mahdollista hyödyntää paitsi CC0:lla lisensoitua metadatan myös linkitettyjä resursseja eli Finnan kautta saatavilla olevia digitaalisia aineistoja niille määriteltyjen käyttöoikeuksien mukaan. Näin on menetelty esim. oppimisympäristöpiloteissa, joissa on tehty ensimmäisiä toteutuksia Finnan kautta saatavilla olevien tietojen ja aineistojen hyödyntämisestä opetukseen tarkoitetuissa sovelluksissa. Tämän tyyppisessä käytössä tulee todennäköisesti korostumaan tarve koneluettavalle oikeuksien hallintaan liittyvälle metatiedolle.

Tarpeita oikeuksien hallintaan liittyville metatiedoille Finnan näkökulmasta

Finnan ja muiden metadatan aggregointiin perustuvien palveluiden kuten Melindan tai Etsimen kannalta olisi tärkeää pyrkiä mahdollisimman standardoituihin käytäntöihin oikeuksien hallintaan liittyviä metatietoja koskien. Tarvitaan kuvailusääntöjen, metatietoformaattien ja tietojärjestelmien (Finnan näkökulmasta taustajärjestelmien) määrätietoista kehittämistä. Myös mahdollisimman laajaa semanttista yhteentoimivuutta tulisi tavoitella, vaikka eri toimijoiden ja toimialojen tarpeet ja käytänteet eroavaisivatkin toisistaan. Samoin tulisi pitää silmällä asiaan liittyvää kansainvälistä kehitystä. Jos tällaisia kehitystoimia ei toteuteta, erilaisia oikeuksien hallintaan liittyviin metatietoihin nojaavia ominaisuuksia ja toiminnallisuksia on hyvin vaikeaa toteuttaa loppukäyttäjille tarkoitettuihin palveluihin. Myös metatiedon ja aineistojen (jatko)käyttömahdollisuudet vaarantuvat.

Esimerkiksi Finnan jo melko yksinkertaiseen, tällä hetkellä pitkälle CC-lisensseihin nojaavan Sallitut käyttötavat -hakurajaimen toteutukseen liittyi haasteita, koska saman lisenssin nimikin voidaan kuvata eri organisaatioiden tuottamassa metadatatassa eri tavoin, puhumattakaan lisensseihin liittyvistä muista tiedoista. Tarve yhtenäistettyihin ratkaisuihin ja kuvailukäytäntöihin on siis Finnan näkökulmasta hyvin ilmeinen. Toki ongelmia tulee jo yksittäisessä tietokannassakin, jos lisenssien kuvailu on kirjavaa.

Lisenssin nimen lisäksi olisi välttämätöntä tallentaa myös linkki (URI) lisenssin tiivistelmään tai kokotekstiin. Linkkitietoa tarvitaan sekä loppukäyttäjäpalveluiden että laajempien aggregoinnin tarpeiden näkökulmasta. Formaattit eivät ole ongelma, koska linkkitieto on tallennettavissa useisiin olemassa oleviin metatietoformaatteihin, kuten kirjastojen MARC 21:een. Finna ei vielä hyödynnä näitä linkkitietoja, koska niitä ei vielä Finnaan juurikaan saada. Linkit lisenssien

¹⁵ Finna-wikin Ohjeita aineistojen käyttöoikeustiedoista -sivulla on lisätietoa metatiedosta, jota Sallitut käyttötavat -rajain tukee, <https://www.kivi.fi/x/iQQeAw>

tiivistelmiin tai kokoteksteihin voidaan kuitenkin luoda Finnassa metadatas-
olevaan lisenssimerkintään perustuen. Jatkossa, kun tallennettu linkkitieto
yleistyy, on tarkoituksenmukaista, että Finna ja vastaavat sovellukset osaavat
sitä hyödyntää ja esittää linkit asianmukaisella tavalla.

Oikeuksien hallintaan liittyvien metatietojen tulisi Finnan näkökulmasta olla
mahdollisuuksien mukaan sekä koneluettavia että ihmisille ymmärrettäviä.
Molempia tarpeita palvelisi metatietojen, kuten yksinkertaisimmillaan lisenssien
nimien kirjoitusasujen, yhteismitallisuus. Ihmisymmärrettävyyteen liittyviä
tarpeita eri käyttäjäryhmien kannalta olisi Finnassa jatkossa hyödyllistä selvittää
käytettävyydestein.

CC-lisenssit ja Public Domain Mark ovat Finnan ja muiden
tiedonhakujärjestelmien kannalta selkeitä ja helposti hyödynnettävissä olevia
tapoja ilmaista käyttöoikeuksia. Pelkkä lisenssin nimi ja muut lisenssiin liittyvät
tiedot eivät kuitenkaan kata kaikkia käyttöoikeustietoihin liittyviä tarpeita, ja tätä
varten olisi tärkeää olla perillä käyttöoikeustietoihin liittyvästä kansainvälisestä
kehityksestä. Esimerkiksi Finnassa saatavilla olevien kulttuuriperintöaineistojen
kannalta on hyvä seurata Europeanassa tapahtuvaa kehitystä sekä European ja
Digital Public Library of American (DPLA) RightsStatements.org -yhteistyötä, sillä
Finnaan haravoidaan joka tapauksessa myös esimerkiksi Europeanan päätyvien
aineistojen metatietoja.

Tutkimusaineistojen arkkitehtuurit

Tutkimusaineistolla tarkoitetaan aineistoa joka syntyy tai jota käytetään
tutkimusprosessin aikana, jolloin tutkija käytännössä useimmiten toimii jossakin
tutkimusorganisaatiossa. Tutkimusaineistojen hallinta on haasteellista, koska se
vaatii paljon substanssiosaamista, mutta usein myös tiedonhallinnallista ja
juridista osaamista. Perinteisesti varsinkin yliopistoissa tutkijoilla on ollut varsin
suuri henkilökohtainen sananvalta ja eettinen vastuu. Sähköiset aineistot ovat
kuitenkin muuttaneet tilannetta niiden hallinnan vaikeuden vuoksi. Tässä luvussa
esitellään ensin tutkimusorganisaatioiden linjauksia oikeuksien suhteen,
muutamia tärkeitä palveluita sekä muiden tuottamia tutkimuksessa käytettäviä
aineistoja, jotka ovat henkilötietojen vuoksi vaikeita tutkimusaineistoina sekä
joitakin muita esimerkkiaineistoja. Kuvaus ei ole millään tavalla kattava
(varsinkaan jälkimmäisten kohdalla), mutta olemme pyrkineet nostamaan esiin
eri tyyppisiä aineistoja ja tarpeita. Ongelmat juontuvat usein epäselvistä tai
puuttuvista sopimuksista tai laintulkinnan hankaluuksista. Tutkimuskäyttöön
liittyy myös erilaiset tarpeet tutkimuslupiin sekä niihin liittyvien vastuiden
hallinnointiin. Tähän löytyy malli genomitieteestä, jossa on luotu Data Access
Committee -järjestelmä.¹⁶

Tutkimusorganisaatiot

Tutkimusorganisaatioilla tarkoitetaan tässä yliopistoja, korkeakouluja ja
tutkimuslaitoksia. Näissä on viime vuosien aikana kirjoitettu datapoliittisia
linjapapereita, joiden avulla pyritään luomaan yhtenäisempiä käytäntöjä
tutkimusaineistojen suhteen. Yliopistot muodostavat siinä mielessä
poikkeuksellisen kokonaisuuden, että tutkijoiden omistajuus, vastuu ja kontrolli
tutkimustuotostensa suhteen on usein ollut vahva ja itsenäinen verrattuna
esimerkiksi yritysten tutkimus- ja kehitystoimintaan. Tämän taustalta löytyvät
sekä historialliset syyt, kuten humanististen aineiden tuotosten tekijänoikeuden
kannalta selkeä asema teoksina, sekä ideologiset sivistykseen ja luovuuden
tukemiseen liittyvät arvot, jotka perustuvat ajatukseen akateemisen tutkimuksen
vapauden tärkeyteen tieteellisen tietämyksen kasvun laadun ja edistyksen
takeena.

¹⁶ Ks. Data Access Committee EGA.

https://www.ebi.ac.uk/ega/submission/tools/EGA_webin#Webin_DAC

Tutkimusjulkaisuja kerätään monessa organisaatiossa julkaisuarkistoon. Suomalaisilla yliopistoilla OpenAIRE-yhteentoimivia julkaisuarkistoja on tällä hetkellä kuusi kappaletta: Aaltodoc, Helda, JyX, LUTPub, Theseus ja Jultika, lisäksi joillakin tutkimuslaitoksilla on omansa. Yliopistojen open access- politiikan mukaisesti julkaisut tallennetaan näihin avoimiin julkaisuarkistoihin. Näiden julkaisuarkistojen metatiedot haravoidaan automaattisesti OpenAIRE-portaaliin. Myös Zenodoon tallennetut julkaisut haravoidaan OpenAIREen ja Zenodoa suositellaankin ainakin Aalto yliopistossa tutkijoille, joiden oma organisaatio ei vielä tarjoa OpenAIRE-yhteentoimivaa julkaisuarkistoa. Myös Zenodoon tallennetut tutkimusaineistojen metadatat haravoidaan OpenAIRE portaaliin.

Paljon tietoa kertyy myös tutkimustietojärjestelmiin, joita myös kutsutaan CRIS-lyhenteellä (Current Research Information System). Yksittäinen yliopisto ei välttämättä voi pyytää ostamaansa omaan tutkimustietojärjestelmään uutta toiminnallisuutta tai ominaisuutta, vaan kaikki kehittämissuunnitelmat pyritään esimerkiksi Pure-järjestelmien kohdalla tekemään kansallisen käyttäjäryhmän kautta. Yhteisiin sopimuksiin pääseminen määrittelyissä on ollut haastavaa, koska Puren käyttäjinä on eri kokoisia ja eri tieteenaloja edustavia yliopistoja, joilla on hyvinkin erilaiset julkaisukäytännöt ja tarpeet. Kansainvälisesti yhteensopivien standardien ja käytäntöjen suosiminen kansallisella tasolla helpottaa tietojärjestelmien ylläpitoa ja määrittelyjen ylläpitoa tulevina vuosina ja varmistaisi yhteentoimivuuden. Muutokset määrittelyissä teettävät paljon työtä tietojärjestelmien ylläpitäjille, koska ne vaikuttavat usein moniin eri elementteihin järjestelmissä (tallennuslomakkeet, TENK CV-pohja, luokitukset, portaali, raportointi, tietovarastomäärittelyt jne). Tietomäärittelytyötä tehdään myös Tutkimuksen tuen ja hallinnon -verkostossa.¹⁷

Sähköisten tutkimusaineistojen hallinta on haasteellista ja vaatii niiden koko elinkaaren hallintaa. Tutkimusorganisaatioiden tavoitteet ottaa tutkimusaineistot haltuunsa ja turvata niiden laatu ja asianmukainen säilytys voi kuitenkin olla osittain ristiriidassa tutkijoiden vapauden ja kontrollin tarpeen ja siihen liittyvän vastuun kanssa. Jos asiaa lähestytään prosessi- ja palvelulähtöisesti, nämä ongelmat lienevät useimmiten ratkaistavissa, kunhan palvelutarjonta on riittävää. Yliopistoissa datapolitiikka-aihetta on lähestytty hieman eri tavoin. Avoimuus on kaikille yhteinen ja kaikkien kannalta rationaalinen tavoite, mutta toisaalta ymmärretään hyvin joihinkin aineistoihin liittyvä rajoitusten tarve. Aineistojen hallintaan käytettävät järjestelmät voivat olla joko julkaisuarkistoja, tutkimustiedonhallintajärjestelmiä tai ulkoisia palveluita. Metatietoja tuotetaan eri järjestelmissä eri tavoin. Tässä käsitellään keskeisimpiä datapolitiikoiden oikeuksiin liittyviä linjauksia.

Aalto-yliopisto julkaisi tutkimusdatan hallintapolitiikkansa helmikuussa 2016. Sen mukaan tekijänoikeuden suojaaman tutkimusdatan omistajuus siirtyy oikeuksiensiirtosopimuksen nojalla Aalto-yliopistolle niissäkin tapauksissa, joissa data on luotu ulkoisen rahoituksen projektissa. "Tutkimusryhmän vastuullinen johtaja tekee tämän datan avaamista koskevat päätökset. Päätöksenteossa noudatetaan eettisiä periaatteita ja hyvää tieteellistä käytäntöä, hankesopimuksia sekä ohjeistusta henkilötietojen käsittelystä tieteellisessä tutkimuksessa. [...] Aalto-yliopiston tutkimusdatan hallintapolitiikka koskee tutkimushankkeissa tuotettavia, käytettäviä ja muokattavia digitaalisia tutkimusaineistoja eli tutkimusdataa." Lisäksi mainitaan, että tutkimusdatalle "voidaan sopia embargoaika tukemaan tutkimusdatan luojan strategisia tavoitteita. Embargoaika ilmoitetaan tutkimusdatan yhteydessä."

¹⁷ Tiedot tutkimustieto- eli CRIS-järjestelmistä tulevat Aalto yliopiston kommentaareista tämän selvityksen versioon 0.9. TUHA-verkoston työtä dokumentoidaan sen wikisivuilla.
<https://confluence.csc.fi/pages/viewpage.action?pageId=39061987>

Aalto-yliopiston linjauksissa painottuu strateginen avaaminen, avoimia lisenssejä kuten CC suositellaan, mutta huomioidaan tilanteet, joissa datan omistajuus ei ole siirtynyt yliopistolle tai sen julkaiseminen ilman rajoituksia ei ole mahdollista strategisista syistä.¹⁸

Helsingin yliopiston tutkimusdatapolitiikka hyväksyttiin noin helmikuussa 2015, ja se on toiminut innoittajana ja esimerkkinä monelle toimijalle. Helsingin yliopisto ilmoittaa datapolitiikassaan tarjoavansa tukea tutkimusaineistoihin liittyvien oikeudellisten kysymysten tunnistamiseen ja ratkaisemiseen: "Vastuulliset tutkijat vastaavat siitä, että tutkimuksessa syntyvien aineistojen omistus- ja käyttöoikeuksia koskevat sopimukset tehdään mahdollisimman varhaisessa vaiheessa, soveltuvin osin ennen tutkimushankkeen alkua." Lisäksi linjataan, että yliopiston piirissä tuotettu, julkaistuihin tutkimustuloksiin liitetty tutkimusdata on lähtökohtaisesti yhteiskäyttöistä ja avointa.¹⁹ Helsingin yliopisto lausuu myös selkeästi, että tutkijoiden on kannettava vastuu aineistojensa tietoturvasta, eettisyydestä ja laillisuudesta.²⁰ Tämä on johdonmukainen seuraus tutkijoiden vapaudesta ja kontrollista omiin tuotuksiinsa.

Itä-Suomen yliopiston julkaisu- ja datapolitiikassa joka julkistettiin joulukuussa 2015 näkyy muun muassa terveystieteiden vahva asema yliopistossa. Eettisyys ja tietoturva ovat tärkeässä roolissa. Aineistojen käyttö- ja omistusoikeuksista on ohjeen mukaan sovittava ennen tutkimuksen aloittamista. Kohdassa neljä puhutaan erikseen sopimustutkimuksesta: "Yliopistossa tehdyn tutkimuksen, erityisesti sopimustutkimuksessa syntyneet tutkimustulokset kuuluvat yliopistolle siinä laajuudessa kuin lainsäädäntö; kutakin sopimustutkimushanketta koskevat rahoitusehdot ja/tai kutakin sopimustutkimushanketta koskevat sopimukset sitä edellyttävät. Sopimustutkimusta on mm. maksullinen palvelutoiminta sekä tutkimus, jossa on mukana ainakin yksi yliopiston ulkopuolinen taho (osatutkimuksen tekijänä, rahoittajana tai muuna osallistujana) ja johon liittyy tutkimuksen tuloksia tai sen toteuttamistapaa koskevia velvollisuuksia (Laki oikeudesta korkeakouluissa tehtäviin keksintöihin 369/2006). Avoimessa tutkimuksessa aikaansaadut tulokset kuuluvat lähtökohtaisesti tekijälle." Avoimuus on lähtökohta aina kun se mahdollista. Kohdassa yhdeksän sanotaan, että jos "tutkimusaineistoihin liittyy kolmansien osapuolten oikeuksia, tutkimusaineistoja ei lähtökohtaisesti saateta avoimeen käyttöön. Potilastutkimuksiin liittyvät tutkimusaineistot ovat pääsääntöisesti arkaluonteisia, jonka johdosta niitä ei saateta avoimeen käyttöön." Lisäksi vaaditaan yliopiston mainitsemista lähteenä kun aineistoja käytetään. Yliopisto pitää myös itsellään oikeuden periä korvauksia esimerkiksi elinkeinoelämän toimijoilta.²¹

Jyväskylän yliopiston rehtori teki päätöksen tutkimuksen tietoaineistojen hallinnan periaatteista jo vuonna 2014.²² Linjauksissa painotetaan toisaalta eettisyyttä, toisaalta avoimuutta. Toimenpiteenä mainitaan, että tutkimusryhmän johtaja sopii yhdessä ryhmän tutkijoiden kanssa, mitä aineistoja tulee tallentaa ja avata muille tutkijoille. Ilman ryhmää toimiva itsenäinen tutkija sopii aineistojen hallinnasta laitoksen johtajan kanssa ja väitöskirjatyöntekijä ohjaajansa kanssa. Omistus- ja käyttöoikeuksista tulee sopia jo aineistonhallintasuunnitelmaa laadittaessa. Mahdollisina osapuolina mainitaan yliopisto ja sen yksiköt, yhteistyöorganisaatiot, rahoittajat, tutkimusryhmät, tutkijat ja ihmistieteissä myös tutkimuksen kohteet. Yliopistolla käytössä olevassa

¹⁸ Aalto-yliopiston tutkimusdatan hallintapolitiikka. Rehtorin päätös 10.2.2016
<http://www.aalto.fi/fi/midcom-serveattachmentquid-1e5ec2a8641fea4ec2a11e5b740c53cbf4c62406240/datapolitiikka.pdf>

¹⁹ Helsingin yliopiston datapolitiikka
<http://www.helsinki.fi/kirjasto/fi/avuksi/tutkimusdatan-hallinta/tutkimusdatapolitiikka/>

²⁰ Oikeudelliset kysymykset
<http://www.helsinki.fi/kirjasto/fi/avuksi/tutkimusdatan-hallinta/oikeudelliset-kysymykset/>

²¹ Itä-Suomen yliopiston Julkaisu- ja datapolitiikka. 1.12.2015.
<http://www.uef.fi/documents/325783/335637/UEF+julkaisu+ja+datapolitiikka/00266e55-b38b-48c3-be6b-f98550643002>

²² Tutkimuksen tietoaineistojen hallinnan periaatteet Jyväskylä yliopistossa. Rehtorin päätös 9.9.2014
https://www.jyu.fi/hallinto/strategia/periaatteet/tutkimuksen_tietoaineistojen_hallinnan_periaatteet

oikeuksiensierotusopimuksessa sovitaan tulosaineistojen lisäksi myös tutkimusaineistojen omistus- ja käyttöoikeuksista.

Oulun yliopisto on ottanut oman tutkimusdatapolitiikkansa perustaksi Iso-Britannian Concordat on Open Research Data -sopimuksen²³. Datapolitiikassa painotetaan avoimuutta ja hyvää datanhallintaa. Kuudennessa periaatekohdassa puhutaan tekijän- tai tutkijanoikeuksista: "Tutkimusaineiston perustajien oikeudet sen ensisijaiseen käyttöön täytyy tunnustaa". Viides periaate linjaa, että "On olemassa selkeitä perusteita, miksi tutkimusaineistojen avaamista on tarpeen rajoittaa, mutta näiden perusteiden täytyy olla oikeutettuja ja toteennäytettävissä.." ²⁴

Tampereen teknillisen yliopiston tutkimusdatapolitiikka hyväksyttiin tammikuussa 2016 ja se painottaa jokaisen yliopistoyhteisön jäsenen eettistä vastuuta tutkimusaineistojen hallinnassa. Yliopisto tarjoaa tukea tausta- ja tutkimusaineistoihin liittyvien oikeudellisten kysymysten tunnistamiseen ja ratkaisemiseen sekä niitä koskevien sopimusten neuvottelemiseen. Täälläkin vastuulliset tutkijat vastaavat siitä, että tutkimuksessa käytettävien ja syntyvien aineistojen omistus- ja käyttöoikeuksia koskevat sopimukset tehdään mahdollisimman varhaisessa vaiheessa niin, että tutkimustyö turvataan. Yliopiston piirissä tuotettu, julkaistuihin tutkimustuloksiin liittyvä tutkimusaineisto on lähtökohtaisesti yhteiskäyttöistä ja avointa. ²⁵

Turun yliopiston datapolitiikka hyväksyttiin rehtorin päätöksellä helmikuussa 2016. Sen ulkopuolelle on rajattu fyysiset ja biologiset aineistot ja sen johtavana teemana on avoimuus. Myös Turun yliopisto pitää itsellään oikeuden peria korvauksia aineistoista. Yliopisto vaatii aineistojen omistajan ja oikeudellisten rajoitusten mainitsemista kuvailutiedoissa. Aineiston luoja suodaan ensikäyttöoikeus. Lähtökohta on, että vähintään tutkimusaineistoihin liittyvät käyttöoikeudet ovat yliopistolla. Aineistojen käyttöoikeudet on kuitenkin selvitettävä tapauskohtaisesti ja tutkijoiden vastuuta alleviivataan. ²⁶

Tutkimuslaitoksista THL on julkaissut datapolitiikan (ks. alla). Tavoitteina mainitaan käyttöluvien ja tietoluovutusten kansallinen yhden luukun palvelu sekä edistävät saatavuutta laatimalla luovutusten ja omistajuuden pelisäännöt. ²⁷

Myös ympäristöntutkimuksen alalla toimivien tutkimuslaitosten LYNET-verkosto on tehnyt oman aineistopolitiikkansa, jota on päivitetty vuonna 2012. ²⁸ LYNET-laitokset tarjoavat useita satoja luonnonvara- ja ympäristöaineistoja vapaasti hyödynnettäväksi. Laitokset pitävät kuitenkin itsellään tekijänoikeuden ja aineiston alkuperän mainitsemista edellytetään käyttäjiltä. Myös tutkijoiden ensikäyttöoikeus heidän keräämiinsä aineistoihin tunnustetaan.

Vain osassa linjauksissa esiintyy hieman ongelmallinen omistajuus-käsite tai -teema. Tutkijoiden vastuuta painotetaan sopimusten tekemisessä ja organisaatioiden vastuuta tarjota tukea ja palveluita. Pitemmän päälle ei kuitenkaan ole toimivaa, että tutkijat säilyttävät hyvin pitkän tai pysyvän päätösvallan ja kontrollin aineistoihinsa ja järjestelmiä suunniteltaessa ja sopimuksia tehtäessä onkin tarkoituksenmukaista huomioida koko

²³ Concordat on Open Research Data

<http://drupal oulu.fi/yliopisto/sites/default/files/content/ConcordatOpenResearchData.pdf>

²⁴ Avoin tiede ja tutkimus. 5.2.2016. Oulun yliopisto.

<http://www oulu.fi/yliopisto/tutkimus/tutkimuksen-tuki-tutkijoille/avoin-tiede-ja-tutkimus>

²⁵ TTY:n tutkimusdatapolitiikka [Rehtorin päätös 27.1.2016]

<http://scienceport.tut.fi/c.php?q=217025&p=2944899>

²⁶ Turun yliopiston datapolitiikka 2016. [Rehtorin päätös 9.2.2016]

https://www.utu.fi/fi/Tutkimus/avoin-tiede/tutkimusaineistot/Documents/ty_datapolitiikka.pdf

²⁷ Datapolitiikka. THL. <https://www.thl.fi/fi/tutkimus-ja-asiantuntijatyo/datapolitiikka>

²⁸ LYNET organisaatioiden aineistopolitiikka.

http://www.lynet.fi/fi-FI/Ajankohtaista/Uutiset/Paivitetty_aineistopolitiikka_LYNETlaito%2838320%29

tutkimusprosessi, aineistojen elinkaari sekä sen hallintaan liittyvät eri vaiheet ja muutostarpeet. Muuten suomalaisissa datapolitiikoissa on huomioita niin käyttörajoituksiin, embargoihin kuin maksullisiin aineistoihin liittyen. Ominaista tutkimusaineistoille on myös sopimusten suuri määrä eri osapuolten välillä. Missä määrin niitä voi ja on mielekästä standardoida ja saattaa koneluettavaan ja -ymmärrettävään muotoon on vaikeaa arvioida.

Tieteenalariippumattomat tutkimusaineistopalvelut

Etsin

[Etsin](#) tutkimusaineistojen hakupalvelu on Avoin tiede ja tutkimus -hankkeessa toteutettu tietovaranto ja palvelu, joka on ollut tuotannossa joulukuusta 2014, aineistoja lähes 10 000 (huhtikuussa 2016). Aineistojen metatietoja voi kirjautumatta hakea verkkoliittymän kautta. Aineistojen kuvailua voi tehdä kirjautuneena suoraan Etsimen tietovarantoon. Palveluun haravoidaan myös tutkimusaineistojen metatietoja muualta (Tietoarkisto, Kielipankki). Etsin liittyy myös suoraan tutkimusaineistojen pitkäaikaissaatavuuden palvelukokonaisuuteen sekä aineistojen verkkojulkaisupalveluun AVAAseen.

Etsimen tietomalli pohjautuu TTA-hankkeen minimimetatietomallille²⁹, jossa on kenttä toimija ja sillä roolit tekijä, muu tekijä, tuottajaprojekti, jakelija/julkaisija, omistaja ja rahoittaja. Tällä hetkellä niihin kytkeytyy muutamia oikeuksia. Oikeudet metatietojen editointiin on tällä hetkellä sillä käyttäjällä, joka on vienyt tiedot tietokantaan mutta niitä voi lisätä. Jakelijalla on oikeus jakaa samassa palvelukokonaisuudessa olevia käyttörajoitettuja aineistoja REMS-järjestelmän avulla. Lisäksi aineisto on linkitettävä organisaatioon. Organisaatioilla ei vielä on editointi- tai muita hallinnointiin liittyviä oikeuksia aineistoihin tai metatietoihin, vaan näistä tulisi sopia. Etsimessä olisi siis tarpeen tapa kuvailla tarkasti mutta helposti oikeuksia tavalla joka vastaa mm. yliopistojen datapolitiikkoja, joista jotkut esimerkiksi vaativat että yliopisto on mainittava tutkimusaineistojen "lähteenä". On tärkeää ottaa huomioon, että metatietoja voivat viedä tietokantaan myös henkilöt, jotka eivät tunne sen paremmin Etsimen tietomallia kuin kuvailun periaatteitakaan. Tallennuksessa sovellettava käyttöliittymä tulee toteuttaa niin, että se tukee metatietojen syöttöä mahdollisimman hyvin, esim. ohjaamalla tallentajaa nimien, asiansanojen, tunnisteiden ja käyttöoikeuksien metatietojen syötössä.

Etsimen tehtävänä on myös jakaa metatietoja avoimena datana esimerkiksi kansainvälisiin palveluihin. Siksi tarvitaan kuvailumalli, jonka on oltava linkitettävissä [DCAT-sovellusprofiiliin](#) Euroopan komission edellyttämällä tavalla.

Tällä hetkellä aineiston kuvailija kuvailee Etsimessä myös käyttöoikeudet manuaalisesti. Kuvailua on tuettu tarjoamalla vaihtoehtoisia koneluettavia lisenssejä. Kuvailija valitsee:

- *Oikean saatavuusluokan*
 - o avoimesti/avoimesti rekisteröitymällä tai tunnistautumalla/lupahakemuksella/ottamalla yhteyttä yhteyshenkilöön
- *Sopivan käyttöluvan* alusvetovalikosta
 - o hyvinmääritellyjä lisenssejä (esim. Creative Commons) sekä vaihtoehdot muu/neuvottelussa/ei määriteltä (Tallennusta ohjaamaan voidaan tehdä [lisenssivalitsin](#))

EUDAT

EU:n rahoittama eurooppalainen EUDAT-palvelu tarjoaa tieteenalariippumattomia palveluita tutkimusdatan hallintaan, jakamiseen ja säilytykseen. Sen

²⁹ TTA-hankkeen minimimetatietomalli:
<http://avointiede.fi/documents/10864/39400/TTA-minimimetatietomalli.pdf/0090c35e-e3fc-40af-8991-ef888e38c890>

hakupalveluun B2FIND voidaan poimia metatietoja OAI-PMH-rajapinnan kautta Dublin Core, ISO 19115 (geodata, ENES), MarcXML, CDMI (CLARIN) sekä DDI -formaateissa. B2FIND kokoaa metadatan useista eri EUDAT-palveluista sekä ulkoisista lähteistä sopimuksen mukaan. B2FIND on CKAN-sovellukseen pohjautuva järjestelmä, jossa ei ole erityisiä ohjeita käyttöoikeuksien metadatan osalta, vaan ainoastaan lisenssitiedon tallentamiseen tarkoitettu kenttä.³⁰

Zenodo ja OpenAIRE

EU:n Horisontti 2020 -hankkeissa edellytetään avointa julkaisemista ja suositellaan tutkimusdatan avaamista. Tämän ehdon toteuttamista tukee OpenAIRE -infrastruktuuri³¹, jota kehitetään keskeisenä osana eurooppalaista avoimen tieteen infrastruktuuria. OpenAIRE-ohjeistus jakaantuu kolmeen osaan sisältäen erilliset ohjeet data-arkistoille, julkaisuarkistoille ja tutkimustietojärjestelmille. Ohjeet data-arkistoille noudattavat pääosin DataCite-organisaation suosituksia³², mutta rights-kentät ovat OpenAIRE-suosituksissa pakollisia silloin, kun niiden käyttö soveltuu aineiston luonteeseen³³. Lisäksi suositellaan rightsURI-kentän käyttöä. Kentän sisällöksi suositellaan esimerkiksi Creative Commons URIa tai oheista info:eu-repo-Access-Terms -sanastoa³⁴, jonka arvoina ovat suljettu, embargossa, rajoitettu ja avoin saatavuus.

Yhteistyössä OpenAIRE-hankkeen kanssa CERN kehittää ja ylläpitää digitaalista Zenodo-arkistoa. Se soveltuu monimuotoisen digitaalisen aineiston tallennukseen joko avoimeen käyttöön tai rajoitetusti esimerkiksi tutkimushankkeen käyttöön. EU:n Horisontti 2020 -rahoitusohjelma ei kehoita käyttämään mitään tiettyä julkaisu- tai data-arkistoa, mutta niille hankkeille, joilla ei ole vakiintunutta julkaisu- ja data-arkistoa, Zenodo tarjoaa käyttökelpoisen vaihtoehdon. Se perustuu CERNin kehittämään Invenio-ohjelmistoon³⁵.

Zenodo-palvelussa metatiedot tallennetaan MARC21-formaattiin. Ne on lisensoitu CC0-lisenssillä lukuunottamatta niihin sisältyviä sähköpostiosoitteita. Tallennettu aineisto voi olla avointa tai suljettua, ja siihen voi liittyä ajastettu julkaisuviive, embargo. Avoimesti julkaistulle aineistolle tallentajan on valittava jokin tarjolla olevista lisensseistä, esimerkiksi CC-lisenssi³⁶. Henkilötietoja voi palvelussa julkaista vain anonymisoituna.

Rajoitetun saatavuuden osalta Zenodo tarjoa muutamaa vaihtoehtoa data-aineiston käytön rajaamiseen:

- Vaatii rekisteröitymistä
- Vaatii lupaa/käyttäjäsopimuksen hyväksymistä (joskus tutkimussuunnitelmaa)
- Ainoastaan paikallinen käyttö
- Vaatii tapauskohtaisen luvan aineiston omistajalta/omistajilta

Aineiston voi myös arkistoida pimeään arkistoon, jolloin siihen ei anneta oikeuksia muille käyttäjille. Aineiston tallentaja voi muuttaa käyttöoikeusehtoja jälkikäteen

Ihmistieteellisten alojen tutkimusaineistot

Ihmistieteissä aineistojen käsittely on usein haasteellista. Tieto voi olla arkaluonteista monesta syystä. Yksi yleinen ja keskeinen peruste on henkilöihin

³⁰ CKAN Metadata <http://ckan.org/features-1/metadata/>

³¹ OpenAIRE Guidelines https://guidelines.openaire.eu/en/latest/data/field_rights.html

³² DataCite schema

https://schema.datacite.org/meta/kernel-3/doc/DataCite-MetadataKernel_v3.1.pdf

³³ Use of DataCite. OpenAIRE.

https://guidelines.openaire.eu/en/latest/data/use_of_datacite.html#accessrights

³⁴ Info EU repo Access Rights <http://purl.org/eu-repo/semantics/#info-eu-repo-AccessRights>

³⁵ Invenio Digital Library Framework. <http://invenio.readthedocs.org/en/latest/index.html>

³⁶ Policies. Zenodo. <https://zenodo.org/policies>

liittyvät tiedot, joita käsitellään tässä yhtenä moniosaisena kokonaisuutena.

Näihin kuuluvat rekisterit ja tutkimuksen yhteydessä käytetyt ja syntyneet tutkimusaineistot. Suomessa on paljon rekistereitä, joista voidaan louhia tietoa esimerkiksi terveydestä, yhteiskunnallisista ja taloudellisista seikoista ja näiden korrelaatioista. Nämä ovat useimmiten salaisia ja erityisten lupamenettelyjen takana olevia aineistoja.

Laillisia ja eettisiä tutkimuksen hidasteita on yritetty ratkaista eri tavoin. Esimerkiksi FMAS (Finnish Microdata Access Services)-palvelu on mainittu maamme tutkimusinfrastruktuurien tiekartassa, joka on Opetus- ja kulttuuriministeriön alaisen Suomen Akatemian linjaus. Tilastokeskus tarjoaa tutkijoille turvallista Fiona-ympäristöä aineistojensa käsittelyyn. Viranomaislupia haetaan nykyään yhteisellä lomakkeella, joka toimitetaan kunkin aineiston omistajalle, joka vastaa osaltaan prosessista. Tärkeitä rekisteritutkimuksen toimijoita on monia. Tunnisteellista tietoa sisältävät kuitenkin paitsi erilaiset viranomaisrekisterit myös esimerkiksi ihmisiä kuvaavat videotallenteet ja heidän puhettaan sisältävät ääninäytteet, joita tutkijat keräävät erilaisiin tutkimustarkoituksiin.

Tietoja jotka tekevät henkilöstä identifioitavan on siis monenlaisia. Suorat tunnisteet ovat yksilöiviä, mutta myös heikommat ja epäsuorat tiedot ovat datan louhinnassa suhteellisen helposti hyödynnettävissä henkilöiden tunnistamiseen. Henkilötietolain mukaan myös pseudonymisoitu aineisto on aina henkilörekisteri. Anonymisoitu aineisto on sellaista, ettei mitään avaimia tai muita mahdollisuuksia rekonstruoida tietoja ole siten, että henkilöt olisivat tunnistettavissa. Tämä tarkoittaa esimerkiksi ettei anonymisoituun aineistoon enää voi lisätä uutta tietoa siellä olevista henkilöistä. Anonymisointi on siis peruuttamaton toimenpide, joka voi heikentää jatkokäytön mahdollisuuksia selyästi. Anonymisoinnista voi lukea lisää Tietoarkiston aineistohallintaoppaasta.³⁷ Anonymisointi on usein edellytys aineiston avaamiselle. Erityisen hankalaa voi olla ääni- ja videoaineistojen anonymisointi. Tutkittavat voivat antaa luvan aineiston julkaisemiseen, mutta eivät avointa lupaa uuteen tutkimuskäyttöön. Tähän löytyy poikkeus vain biopankkilaista niiden sisältämien aineistojen suhteen, joihin liittyy laaja suostumus.

Tietosuojavaltuutetun ohjeiden mukaan henkilötietoja voidaan käsitellä tutkittavien itsensä antamalla suostumuksilla, mikä on ensisijainen tapa, kun tutkimusaineistoon tarvitaan tunnistettavia henkilöitä koskevia tietoja. Aina suostumusten hankkiminen ei kuitenkaan ole mahdollista tietojen suuren määrän, tietojen iän tai muun sellaisen syyn vuoksi. Tällainen syy on usein olemassa, kun tutkimusaineistona halutaan käyttää valtakunnallisten henkilörekisterien sisältämiä tietoja. Tällöin tietojen käyttöön voidaan hakea yllä mainittua viranomaislupaa.

Tunnisteellisten ihmisiin liittyvien aineistojen käyttö vaatii erityistä huolellisuutta. Avoin tiede ja tutkimus -hankkeessa on tehty Arkaluonteisen aineiston viitearkkitehtuuri, jossa on pyritty kuvaamaan kokonaisuutta, jota tutkimus tarvitsee hyödyntääkseen sensitiivisiä aineistoja tehokkaasti, eettisesti ja turvallisesti. Arkkitehtuurissa roolipohjainen käyttö- ja pääsyoikeuksien hallinta on nimetty kriittiseksi käyttöturvallisuuden osaksi (s. 18). Palvelut ja niiden väliset suhteet sekä prosessit on kuvattu arkkitehtuurissa järjestelmästä riippumatta.

Tieteellistä tutkimusta varten muodostettavan henkilörekisterin käytön on aina perustuttava asianmukaiseen tutkimussuunnitelmaan. Tutkimusta varten tutkijoille muodostuu rekisteritutkimuksessa oma, tietolähteenä käytetystä henkilörekisteristä erillinen rekisteri. Useissa tapauksissa luvan myöntämistä

³⁷ Aineistohallinnan käsikirja [verkkójulkaisu]. Tampere: Yhteiskuntatieteellinen tietoarkisto [ylläpitäjä ja tuottaja]. <<http://www.fsd.uta.fi/aineistohallinta/>>. urn:nbn:fi:fsd:V-201504200001 (Viitattu 09.11.2016.)

harkitseva viranomainen pyytää lausunnon tietosuojavaltuutetulta. Tutkimukseen ryhtyvä rekisterinpitäjä vastaa silti aina viime kädessä itse tutkimukseensa liittyvästä henkilötietojen käsittelystä. Käytännössä rekisteriaineistot usein tuhotaan, koska anonymisoitua aineistoa ei ole mielekästä arkistoida, sensitiivisen tutkimusdatan säilyttämiseen sopivaa arkistoa ei ole tai aineiston luovuttaja on vaatinut aineiston tuhoamista.

Tietoarkisto (FSD)

Tietoarkisto on tutkimuksen ja opetuksen valtakunnallinen palveluinfrastrukturi, joka arkistoi ja välittää sähköisiä tutkimusaineistoja tutkimukseen, opetukseen ja opiskeluun. Tietoarkisto edistää tieteen avoimuutta, läpinäkyvyyttä, tiedon kumuloitumista ja olemassa olevan tiedon tehokasta tutkimuskäyttöä.

Tietoarkisto on CESSDAn (Consortium of European Social Science Data Archives) kansallinen palveluntuottaja ja Suomi valtiona on CESSDAn jäsen. CESSDAn jäsenyys sitoo Tietoarkiston toimintaa. CESSDAn keskitettyjen palvelujen rakentaminen on käynnissä. Tällä hetkellä on esimerkiksi suunnitteilla CESSDAn kaikkien palveluntarjoajien yhteinen datan saantipolitiikka.

Tietoarkiston perustehtäviä ovat mm. aineistojen hankkiminen arkistoitavaksi, arkistointi (sisältää mm. aineiston kuvailun) sekä aineistojen toimittaminen jatkokäyttöön. Tietoarkisto tekee aineiston luovuttajan kanssa arkistointisopimuksen aineiston luovuttamisesta Tietoarkistoon. Arkistointisopimuksessa luovuttaja sitoutuu arkistointia ja jatkokäyttöön toimittamista koskeviin ehtoihin. Sopimuksessa luovuttaja valitsee ehdot aineiston toimittamiseen jatkokäyttöön. Vaihtoehdot ovat:

- kaikkien käytettävissä (ei vaadi jatkokäyttäjän rekisteröitymistä)
- tutkimukseen, opetukseen ja opiskeluun (vaati jatkokäyttäjän rekisteröitymisen)
- vain tutkimukseen (vaatii jatkokäyttäjän rekisteröitymisen)
- vain luovuttajan luvalla (vaatii jatkokäyttäjän rekisteröitymisen)

Aineiston luovuttaja voi asettaa sopimuksessa määräaikaisen embargon eli ajan, jolloin aineistoa ei vielä saa toimittaa jatkokäyttöön. Luovuttajan on myös mahdollista kirjata sopimukseen vapaamuotoisia aineistokohtaisia erityisehtoja.

Aineistojen jatkokäyttöön saaminen tapahtuu palveluportaali Ailan³⁸ kautta. Ladatessaan aineiston Ailasta jatkokäyttäjä sitoutuu aineistojen yleisiin käyttöehtoihin³⁹. Niissä sitoudutaan mm. siihen, että käyttöluva on henkilökohtainen; aineisto ei saa olla ulkopuolisten ulottuvilla; aineistoa saa käyttää vain ilmoitettuun käyttötarkoitukseen; aineistoon viitataan asianmukaisesti; julkaisut ilmoitetaan tietoarkistoon jne. Jatkokäyttäjän tulee myös noudattaa aineistoon liittyviä aineistokohtaisia erityisehtoja. Yksi tyypillinen esimerkki erityisehdoista on, että aineistosta tuotetuista julkaisuista tulee lähettää kopio aineiston tekijöille.

Aineiston jatkokäyttäjä sitoutuu myös käyttämään aineistoa vain aineiston käyttöoikeuden mukaisesti (kaikkien käytettävissä; tutkimukseen, opetukseen ja opiskeluun; vain tutkimukseen tai vain luovuttajan luvalla). Tämä tarkoittaa sitä, että esimerkiksi vain tutkimukseen tarkoitettua aineistoa ei saa ladata ja käyttää kandidaatintutkielmaan. Käyttötarkoituksen lisäksi aineiston jatkokäyttäjän tulee myös ilmoittaa hankkeen/tutkimuksen/kurssin nimi ja antaa lyhyt kuvaus.

Kielipankki

Kielipankki on kansallisen FIN-CLARIN-konsortion tarjoama palvelukokonaisuus, joka on suunnattu kaikille kieliaineistoja käyttäville tutkijoille. Kielipankissa on

³⁸ Tietoarkiston palveluportaali Aila <https://services.fsd.uta.fi/>

³⁹ Tietoarkiston aineistojen yleiset käyttöehdot. <https://services.fsd.uta.fi/docs/terms-of-use>

laaja valikoima teksti- ja puheaineistoja, joista voi tehdä monipuolisia hakuja. Aineistoja voi niiden käyttöehdoista riippuen tutkia ja käsitellä virtuaalisessa työtilassa Kielipankin työkaluin tai ladata omalle koneelle. Kielipankkiin voidaan tallentaa myös muihin kuin kielitieteellisiin tutkimustarkoituksiin kerättyjä aineistoja, mikäli ne sisältävät puhuttua tai kirjoitettua kieltä. Hyvien hakutoimintojen vuoksi myös muut humanistit ja esimerkiksi yhteiskuntatieteilijät käyttävät kasvavassa määrin Kielipankin palveluita.

FIN-CLARIN-konsortioon kuuluu joukko suomalaisia yliopistoja ja tutkimusorganisaatioita sekä CSC – Tieteen tietotekniikan keskus. FIN-CLARIN edustaa Suomea eurooppalaisessa CLARIN-tutkimusinfrastruktuurikonsortiossa, jonka jäsen Suomi on. FIN-CLARIN vastaa kieliaineistojen hankinnasta ja vastaanottamisesta sekä kieliaineistojen arkistointiin ja jatkokäyttöön liittyvien palveluiden kehittämisestä. FIN-CLARIN tarjoaa myös kieliaineistojen hyödyntämiseen liittyvää koulutusta.

CLARIN-infrastruktuuriilla on yhteinen lisenssiluokitus, jota myös Kielipankki noudattaa. CLARIN-lisenssijärjestelmässä lisenssit on jaettu kolmeen pääluokkaan: julkinen pääsy (PUB), pääsy luotetun verkoston jäsenille (ACA) sekä rajoitettu pääsy (RES). Kuhunkin pääluokkaan voi liittyä lisäehtoja, jotka on jaoteltu seuraavasti:

- Tunnistamiseen ja pääsyyn liittyvät ehdot (ID, AFFIL, PERM, FF, PLAN)
- Yleiset käyttöehdot (BY, NC, INF, LOC, LRT, PRIV)
- Levittämiseen liittyvät ehdot (NORED, ND, DEP, SA)
- Muut käyttöehdot (ts. mahdolliset aineistokohtaiset erityisehdot)

Kieliaineistojen kohdalla ACA-luokitus tarkoittaa, että aineistoon pääsee suoraan kirjautumalla luottamusverkostoon kuuluvan organisaation tunnuksella, mikäli kirjautumisen mukana siirtyy tieto, että henkilöllä on organisaatiossaan tutkimukseen liittyvä työtehtävä. RES-luokkaan kuuluvat aineistot edellyttävät aina sekä kirjautumista että henkilökohtaisesti myönnettävää käyttöoikeutta. Ennen aineiston käyttämistä tai henkilökohtaisen käyttöoikeuden myöntämistä käyttäjän on hyväksyttävä Kielipankin yleiset käyttöehdot sekä mahdolliset aineistokohtaiset käyttöehdot.

Kielipankissa saatavilla olevien RES-tyyppisten aineistojen henkilökohtaista käyttöoikeutta voi hakea Kielipankin oikeudet -palvelussa, joka on Kielipankin tarpeisiin räätälöity CSC:n tuottaman REMS-palvelun instanssi. Palvelun kautta voi hakea oikeuden myös Kielipankin ACA-aineistoihin, jos pääsyä ei ole voitu suoraan myöntää pelkällä kirjautumisella (esim. jos oma työtehtävä ei liity tutkimukseen tai jos hakija on opiskelija).

Aineiston käyttöoikeusluokka ja tarkemmat lisenssiehdot määritellään tallennussopimuksessa, jonka FIN-CLARIN tekee aineiston luovuttajan kanssa. Kielipankin aineistoja koskevat lähinnä tekijänoikeuteen ja henkilötietoihin liittyvät käyttörajoitukset. Perusteita löytyy tekijänoikeuslaista, WIPOn suosituksista ja Bernin konventiosta, henkilötietolaista, julkisuuslaista ja arkistolaista.

Tekstiä sisältäviin aineistoihin liittyy tyypillisesti tekijänoikeuksiin pohjautuvia käyttörajoituksia. Tekstien tutkimus- ja opetuskäyttö on tietyissä rajoissa mahdollista yksittäisille tutkijoille ns. digiluvan perusteella. Aineistot saa esim. tallentaa suljettuun sisäverkkoon oman opetus- tai tutkimusryhmän saataville kurssin tai muun opintokokonaisuuden ajaksi. Kopioidun aineiston voi myös jakaa opetus- tai tutkimusryhmälle sähköpostilla. Sekä Tietoarkisto että FIN-CLARIN ovat pyrkineet auttamaan tutkijoita sopimalla isojen tekstimassojen lisensoinnista kootusti joko suoraan oikeudenhaltijoiden tai niitä edustavan Kopioisto ry:n kanssa.

Kielipankin kautta voidaan jakaa myös ääntä ja videota sisältäviä puheaineistoja sekä muita kieliaineistoja, jotka sisältävät henkilötietoja. Tällaisten aineistojen käyttörajoitukset pohjautuvat suureksi osaksi tutkittavien, haastateltavien tai muiden osallistujien antamaan suostumukseen, jonka aineiston tuottaja hankkii aineiston keräämisen yhteydessä. Suostumuksen yksityiskohdat vaikuttavat mahdollisuuteen välittää aineistoa Kielipankissa sekä aineiston lisenssiin.

Kieliaineistoihin liittyy jatkokäytön kannalta se erityispiirre, että niihin on tapana tutkimuksen kuluessa lisätä ns. annotaatiota joko käsin tai koneellisesti. Annotaatio tarkoittaa dokumenttien osasiin, esimerkiksi yksittäisiin sanoihin tai tiettyihin puhenäytteen ajanjaksoihin liittyvää systemaattista merkkeä. Annotaatio mahdollistaa esimerkiksi monipuoliset ja automaattiset haut ja muodostaa olennaisen osan käytettävästä tutkimusmenetelmästä. Myös annotoitu aineisto pitäisi voida luovuttaa edelleen muiden tutkijoiden käyttöön, jotta tutkimus on toistettavissa ja jotta tutkijat voivat parannella toistensa tuottamaa annotaatiota. Näin samasta aineistosta syntyy jatkuvasti uusia, eri tavoin rikastettuja versioita.

Samasta aineistosta voi tarpeen mukaan olla Kielipankissa useita erilaisia versioita, joihin voi liittyä erilaisia käyttöehtoja ja siten erilainen saatavuus. Monissa tapauksissa esimerkiksi tekstiaineiston sisältämät kokotekstit voidaan tarjota ainoastaan henkilökohtaiseen käyttöön RES-luokan lisenssillä. Samoista teksteistä voidaan kuitenkin näyttää lyhyitä katkelmia myös julkisesti Kielipankin hakupalvelun kautta. Vastaavasti puheaineiston alkuperäiset ääni- tai videonäytteet on yleensä suojattava RES-lisensseillä. Joidenkin puheaineistojen tekstimuotoiset, osittain tai kokonaan tunnisteettomiksi käsitellyt litteraatit voidaan kuitenkin tarjota laajempaan käyttöön joko katkelmina tai kokonaisina tekstidokumentteina. Myös katkelma-aineistoista voidaan tutkia monenlaisia kiinnostavia asioita, mutta kokonaisista dokumenteista koostuvat aineistot ovat välttämättömiä esimerkiksi kielen tilastollisessa mallintamisessa ja tiedonlouhinnassa.

Tällä hetkellä ongelmana on löytää Tietoarkiston ja Kielipankin tunnisteellisten tutkimusaineistojen jatkokäyttöön toimittamisen lailliset perusteet. Molemmille on myönnetty Data Seal of Approval –sertifikaatti eli organisaatiot kykenevät luotettavasti säilyttämään pitkäaikaisesti sähköistä dataa ja mahdollistamaan sen jatkokäytön. Lailliset perusteet tunnisteellisten tutkimusaineistojen toimittamiselle jatkokäyttöön löytyisivät parhaiten, jos Tietoarkistolla ja Kielipankilla olisi lainsäädännöllinen arkistostatus tai toiminta turvattaisiin muilla lainsäädännöllisillä toimenpiteillä, joissa huomioidaan myös EU:n uusi tietosuojasetus.

Sensitiiviset henkilötietoaineistot

Kliininen tutkimus vaatii aina ennakkoilmoituksen Valtakunnalliselle lääketieteelliselle tutkimuseettiselle toimikunnalle, joka tarvittaessa siirtää sen sopivan eettisen toimikunnan käsiteltäväksi. Lääketieteellinen tutkimus, jolla puututaan henkilön koskemattomuuteen, vaatii aina eettisen toimikunnan lupaa. Henkilörekistereitä voi saada tutkimuskäyttöön eri viranomaisilta erilaisten lupaprosessien kautta, mutta ne pyritään aina anonymisoimaan ennen tutkijalle luovuttamista. Jos henkilöiltä kuitenkin kerätään lisää tietoja, tarvitaan tietojen yhdistämiseen tutkittavan suostumus. Onkin tärkeä huomioida tutkimusaineistojen koko elinkaari ja sen eri vaiheissa olevat vaihtelevat vastuut ja oikeudet. Näiden kuvaaminen on haasteellista, mutta mahdollista niissä tapauksissa kun lupaprosessi on selkeä. Ongelmana on toimijarekisterin ja tunnisteiden puute.

KELA - Tietojen luovuttaminen tieteellisiin tutkimuksiin

Kela luovuttaa tietoa vain tieteellistä tutkimusta varten ja edellyttää henkilötietolain (523/1999) ja viranomaisten toiminnan julkisuudessa (621/1999) säatelevän lain noudattamista. Henkilötietoa käsitellessä on aina huolehdittava

siitä, ettei yksittäisten henkilöiden tietoja paljasteta tutkimuksessa. Tietoja ei saa tutkimuksessa käyttää päätöksenteon välineenä.

Kelan tietoja luovutetaan vain luvan saaneille henkilöille. Myös anonymisoidun datan käsitteleminen vaatii salassapitositoumuksen. Mikäli tutkimuksessa käytettyä dataa pystyy tunnistamaan joko suoraan tai välillisesti tai jos koodiavain on säilytetty, katsotaan että tutkimusaineisto sisältää henkilörekisteritietoa.⁴⁰

Tilastokeskus

Myös Tilastokeskus tarjoaa tilastotietojen lisäksi yksikkötason tai ns. mikroaineistoja tieteellisiin tutkimuksiin. Henkilötietoja sisältäviä aineistoja voi, aineistosta riippuen, käyttää joko etäkäyttöpalvelu Fionan kautta tai Tilastokeskuksen tiloissa. Niiden käyttöön saaminen vaatii aina erillisen luvan ja hakijalta edellytetään mm tutkimussuunnitelman esittämistä sekä salassapitosopimusta. Mikäli tutkimuksessa yhdistetään myös muiden aineistoja, yhdistäminen tapahtuu aina Tilastokeskuksessa tilastolain puitteissa. Aineiston kokoamisessa ja luovuttamisessa noudatetaan tilastolainsäädäntöä ja laissa määriteltyjä tietosuojaj- ja salassapitokäytäntöjä.⁴¹

THL

THL:ltä voi hakea omaa tutkimusta varten käyttö lupaa salassa pidettäviin tietoihin. Ennen aineistojen luovuttamista hakijalle, on tehtävä salassapitositoumus, jota ilman aineistoja ei luovuteta. Tietojen luovuttajien kanssa on sovittava muutamasta tärkeästä asiasta jo ennen käyttö luvan hakemista, kuten aineistojen keruutavasta ja -kustannuksista, tietosisällöstä sekä näihin aineistoihin liittyvistä tietoturva-asioiden huolehtimisesta.

Yksittäisten henkilöiden lisäksi myös organisaatiot voivat hakea THL:ltä käyttö lupia. Näissä käyttöoikeussopimuksissa sovitaan myös tutkimustulosten julkaisusta ja omistus- ja käyttöoikeuksista.

THL:n omien tietojärjestelmien käyttöoikeudenhallinnasta on tehty myös sisäisiä määräyksiä. Työtehtävien vaatiessa voi myöntää peruskäyttöoikeuksien lisäksi erityiskäyttöoikeuksia. Hakijan muut mahdolliset voimassa olevat käyttöoikeudet on huomioitava käyttöoikeuksien myöntämisessä mikäli niiden seurauksena muodostuvat epäyhteentoimivat yhdistelmät tai käyttöoikeustasojen päällekkäisyyksiä.

THL:n aineistojen saatavuuteen ja käyttöön vaikuttavat useat lait. THL:n toimintaa sääntelevät lait jotka koskevat viranomaisen laitoksen toimintaa, henkilötietojen käsittelyä, henkilötietorekisterien käsittelyä, terveydenhuollon valtakunnallisten rekisterien käsittelyä sekä lääketieteellistä tutkimustoimintaa.

Lakeja ovat:

- Viranomaisen toiminnan julkisuuslaki:
(<http://www.finlex.fi/fi/laki/ajantasa/1999/19990621>)
- Henkilötietolaki: (<http://www.finlex.fi/fi/laki/ajantasa/1999/19990523>)
- Terveydenhuollon valtakunnallisiin rekistereihin liittyvä laki:
(<http://www.finlex.fi/fi/laki/ajantasa/1989/19890556>)
- Lääketieteellistä tutkimusta koskeva laki:
(<http://www.finlex.fi/fi/laki/ajantasa/1999/19990488>)

THL:n aineiston keruuta ja siirtoa koskevat keskeisimmät lait löytyvät THL:n verkkosivuilta: <https://www.thl.fi/fi/tilastot/tutkimuskaytto/lainsaadanto>

⁴⁰ Kelan ohjeita tietojen luovuttamiseen tutkimustarkoituksiin:

http://www.kela.fi/tutkimusta-tukemassa_tietopyynnot-tieteellisiin-tutkimuksiin

⁴¹ Mikroaineistot. Tilastokeskus. <http://www.tilastokeskus.fi/tup/mikroaineistot/index.html>

Anonymisointitekniikat ja anonymisoinnin purkamiseen tarvittavat toimenpiteet kehittyvät jatkuvasti. Henkilötietoja sisältävät aineistot vaativat anonymisointia. Näiden aineistojen siirtäminen pitkäaikaissäilytykseen on haastavaa, koska anonyyminä pysyminen vuosikymmeniä ei ole taattua. Tekniikkaa kehitty jatkuvasti, joten voidaan siis olettaa että tulevaisuuden tekniikalla salakirjoitukset pystytään helposti purkamaan. Tämä edellyttää että anonymiteetin tasoa tarkistetaan ennen kuin aineistoja voidaan jakaa tulevaisuudessa eteenpäin. Anonymiteetin säilyttäminen on tärkeää henkilötietojen suojan varmistamiseksi.

Toinen haaste aineistojen pitkäaikaisessa säilytyksessä on tiedon tulkitseminen. Arvellaan, että yhteiskunnan muuttuessa, nykypäivän tietoja ei pystytä enää tulevaisuudessa tulkitsemaan. Näin pääsi käymään hoitoilmoitusrekisteri Hilmon pitkien aikasarjojen takia, jolloin sairaalahoito-käsitteen alkuperäinen tarkoitus erosi huomattavasti nykyajan tarkoituksesta. Lisäksi käsitykset aineistojen koosta muuttuvat ajan myötä, niin että nykypäivän suuret aineistot ovat tulevaisuudessa pieniä aineistoja, ja näitä pystytään tulevaisuudessa analysoimaan huomattavasti lyhyemmässä ajassa.

THL:n datapolitiikan tavoitteena on nostaa avoimen datan ja avoimen julkaisemisen toimintansa lähtökohdaksi sekä tehostaa käyttörajoitteisten tietoaineistojen saatavuutta. Tavoitteiden saavuttaminen vaatii käyttöluopaprosessien käsitteleminen läpinäkyvästi yhden luukun palveluperiaatteen mukaisesti, sujuvan yhdistelyn eri viranomaisilta tullutta aineistoa ja aineistojen jakamisen rajapintojen kautta. Jatkokehityksiin kuuluu erilaisten etäkäyttöympäristöjen helpottaminen ja Sitran Palveluoperaattori-aloitteen edistäminen.

Biopankit

Suomessa toimii eri sairaanhoitopiirien, yliopistojen ja Terveiden ja hyvinvoinnin laitoksen ylläpitämiä biopankkeja. Suomen biopankkilaki edellyttää, että nykyiset ja tulevaisuudessa kerättävät tutkimusnäytteet säilytetään biopankeissa, mikäli ne on kerätty biopankkilain tarkoittamalla suostumuksella. Suomessa on tällä hetkellä yhdeksän biopankkia ja suomalaiset toimijat ovat perustaneet oman verkostonsa, BBMRI.fi:n⁴², joka edistää biopankkitoimintaa ja muun muassa yhteentoimivuutta. THL ylläpitää yhtä Suomen biopankeista.

Pohjoismaisen Tryggve-projektin puitteissa etsitään yhteisiä ratkaisuja ja toimintamalleja ihmisten genomidatan ja muun siihen liittyvän sensitiivisen tiedon käsittelyyn. Tryggve on kolmen vuoden projekti, joka tarjoaa yhteistä pohjoismaista alustaa sensitiiviselle datalle. Tryggven avulla sensitiivistä biolääketieteellistä tutkimusaineistoa voi turvallisesti säilyttää, jakaa, analysoida ja suorittaa laskentaa. NeIC (Nordic e-Infrastructure Collaboration) ja ELIXIR:in noodit Norjassa, Ruotsissa, Suomessa ja Tanskassa rahoittavat hanketta. Projektin ydinideana on käyttää pohjoismaissa olemassa olevia palveluita ja voimavaroja. Työt tehdään yhteistyössä käyttäjäyhteisön ja pohjoismaiden tutkimusinfrastruktuuritarjoajien kanssa.^{43 44}

ELIXIR (European life science infrastructure for biological information) on eurooppalainen biotieteisiin erikoistuva tutkimusinfrastruktuuri. Suomen ELIXIR-keskus tarjoaa tutkimusryhmien käyttöön Suomessa kehitettyjä lääketieteen ja bioinformatiikan tietoteknisiä työkaluja. Myös ELIXIRin jäsenorganisaatioiden kehittämät palvelut ja sovellukset ovat suomalaisten tutkijoiden käytettävissä. Suomen ELIXIR-keskus tarjoaa datan tallennuspalveluja, tehokkaita tietokoneita ja työkaluja sekä koulutusta.

⁴² Biobanking and Biomolecular Resources Research Infrastructure <http://www.bbmri.fi/>

⁴³ Nordic e-Infrastructure Collaboration:in sivuilla - Tryggve: <https://wiki.neic.no/wiki/Tryggve>

⁴⁴ NeIC for Open Access to Research Data - esitys: https://www.dropbox.com/sh/2r45n763cda0drk/AABUk1qN_4IrssKdfIHpZmWha/NeIC.pdf?dl=0, sivut 14-18

Biolääketieteen tutkijat pääsevät ELIXIRin kautta käsiksi suureen määrään biologisen tiedon aineistoja.

ELIXIRissä on allekirjoitettava käyttöluopimus ennen tutkimustyön alkua, jossa myöntää luvan muille tutkijoille käyttää omia tietoja tutkimuksellisiin tarkoituksiin. Tämän perusteella tehdään hallintapäätös tai ns. konsentti. Käyttörajoitukset ELIXIRissä perustuvat yksilösuojaan ja EU-direktiiveihin, joiden mukaan tutkimuskäyttö sallitaan, mutta aineistojen jakamista rajoitetaan. Tutkimustuotosten käyttö sallitaan vain tutkimuskäyttöön, ei kaupallisiin tarkoituksiin. ELIXIRin palveluissa tallennettua tutkimusaineistoa voidaan vapaasti käyttää ja jakaa eteenpäin, mutta osalla alkuperäisaineistoilla voi olla kolmansien osapuolten tekemiä käyttöoikeusrajoitteita. Rajoitukset asettavat esimerkiksi patentit, tekijänoikeudet tai muut immateriaalioikeudet.

EGA (European Genome-phenome Archive) on eurooppalainen genomi- ja fenomitiedon arkisto, joka edellyttää että jokainen tutkimustuotos on kytketty johonkin tiettyyn data access-komiteaan. EGAn tutkimuskonsentti sallii tutkimustiedon lähettämisen myös ulkopuolisille tutkijoille, mutta asettaa kuitenkin tietyt rajoitukset datan jakamiseen.

ELIXIRin oma tietokanta, johon on kuvattu alan resursseja. Tietomallissa on mahdollisuus ilmaista resurssille saatavuustieto "Public" tai "Private". Lisäksi voi kertoa lisenssin sekä käyttöehdoista Cost-elementissä arvoilla "Free", "Free with restrictions" tai "Commercial".⁴⁵

Tuoreessa Corbel-hankkeessa mm ELIXIR- ja ECRIN-infrastruktuurit edistävät yhteentoimivuutta biotieteiden alalla laajalla rintamalla. Yksi tavoitteista on saatavuuden harmonisointi.⁴⁶ Kansainvälisesti katsoen Suomen tilanne on hyvä, sillä meillä on luotettava biopankki-verkosto sekä biopankkilain mahdollistamat laajat, selkeät suostumukset. Lisäksi Suomi on osallisena sekä EUDATissa, ELIXIRissä että Corbelissa CSC:n kautta, joten kehitystä on suhteellisen helppo seurata.

Luonnonvara- ja ympäristöaineistot

Luonto- ja ympäristöalan tutkimusaineistoja tuottavat ja arkistovat yliopistotutkijoiden sekä luonnontieteellisten museoiden lisäksi myös tutkimuslaitokset, kuten Suomen ympäristökeskus (SYKE), Luonnonvarakeskus (Luke) ja Elintarviketurvallisuusvirasto Evira. Myös Geologian tutkimuskeskus (GTK), Ilmatieteen laitos (IL), Maanmittauslaitos (MML) ja Teknologian tutkimuskeskus VTT Oy kuuluvat ympäristö ja luonnonvara-alan tietopalveluita tarjoavaan LYNET konsortioon. Näiden lisäksi ympäristöön ja lajistoon liittyviä tutkimus- ja seuranta-aineistoja kerätään myös ELY-keskusten ja kaupunkien ympäristötoimen sekä erilaisten harrastaja- ja luonnonsuojelujärjestöjen puolesta.

Biodiversiteettiaineistot

Luonnontieteellisen keskusmuseon tietoinfrastruktuurin varaan rakentuvan Suomen Lajitietokeskuksen tavoitteena on koota yhteen pilvipalveluun eri organisaatioissa tuotettuja elolliseen ympäristöön liittyviä aineistoja. Aineistot on usein kerätty tiettyä tutkimusta ja tai seuranta varten, mutta avoimina aineistoina niitä voidaan käyttää alkuperäisestä käyttötarkoituksesta paljonkin poikkeavalla tavalla. Tällöin yhdistetyistä aineistoista poimitaan vain tiettyjä lajeja koskevat ja/tai tietyt laatuksiteerit täyttävät tiedot. Uusiokäyttöön poimitun tutkimusaineiston sekaan saattaa päätyä tietoja jopa sadoista alkuperäisaineistoista, mistä syystä aineistokohtaisten rajoitusten ilmaiseminen pelkästään metatiedoissa ei palvele käyttäjää kovin hyvin.

⁴⁵ Bio Tools <https://bio.tools/resources/biotools-1.4/docs/biotools-1.4.html>

⁴⁶ Corbel Project Work Packages <http://www.corbel-project.eu/work-packages.html#c19>

Luonnon monimuotoisuutta eli biodiversiteettiä ja lajien levinneisyyttä kuvaavat aineistot kootaan eri tietolähteistä ja yhdistetään ns. data-matriisiksi, jonka riveinä on yksittäinen havainto lajista tietyssä ajan hetkenä tietyssä paikassa. Näiden tietojen lisäksi havaintoon voi liittyä muuta dokumentaatiota, kuten mittaustuloksia tai faktoja havaitusta yksilöstä tai vallinneista olosuhteista, tai laboratorio analyysyjä näytteeksi kerätystä lajiyksilöstä. Yhteiskäyttöön sopivaa datasta tulee vasta, jos haettavat tietueet matriisin riveillä tulevat yhteisistä auktoriteettitietovarannoista, kuten MML:n maastotietokannasta tai lain mukaisesti ylläpidetystä henkilörekisteristä. Myös lajinimiä pyritään ylläpitämään keskitetyssä nimistötietokannassa.

Keskeisin hakukriteeri data-matriisissa on siis laji, jota ilmaistaan tieteellisiin tarkoituksiin kerätyistä aineistoissa yleensä tieteellisellä nimellä, kun taas Suomen lainsäädännössä lajinimet ilmaistaan ns. kansankielisillä nimillä. Lajeja (ilman synonyyminimiä) on Suomessakin valtava määrä. Suomen lajistoon *arvioidaan* kuuluvan n. 50 000 lajia ja maailman lajilukumäärän arviot vaihtelevat 1,3:sta kahdeksaan miljoonaan. Edes Suomen kaikille lajeille ole mahdollista antaa yleiskielistä nimiä suomeksi tai ruotsiksi. Myös tieteellisten nimien muutoksia ja synonyymejä on lähes mahdotonta hallita, sillä museokokoelmissa käytetyt nimet viittaavat yleensä hyvin vanhaan kirjallisuuteen, minkä lisäksi kokoelmien lajisto sekä tiedonvaihto ovat maailmanlaajuista. Nimistötietokantaan pyritään keräämään mahdollisimman kattava kokoelma kotimaisissa tietokannoissa ja tieteellisessä kirjallisuudessa käytetyistä maailman lajinimistöistä, synonyymeistä ja nimistön hierarkiasta eli taksonomista. Suomessa käytössä oleva nimistö pyritään liittämään maailmanlaajuiseen nimistöön, mutta tällä hetkellä tällaista ei vielä ole käytettävissä. Koska rajoitukset tiedon avoimuuteen koskee yleensä nimenomaan tiettyjä lajia (eikä koko aineistoa), nimistötietokannassa ylläpidetään tietoa siitä mitkä säädökset koskevat kutakin lajia Suomen lainsäädännössä.

Rajoitukset aineistojen avoimuuteen eivät siis liity pelkästään tekijänoikeus- tai henkilörekisterilakiin, vaan myös luonnonsuojelua, vieraslajeja, riistanhoitoa, maatalouden tuhoeläimiä ja metsävaroja koskevassa lainsäädännössä on perusteita tiedonsaannin rajoittamiselle. Rajoitus liittyy usein yksittäiseen lajiin, tai sen sijaintiin lainsäädännön piirissä olevalla alueella. Lainsäädännön perusteella ei kuitenkaan oleteta rajoitettavan koko aineiston saatavuutta, vaan yleensä selvittää karkeistamalla aineistosta poimittavissa olevien yksittäisten lajien tietoja.

Yksittäistä aineistoa koskevissa metatiedoissa tulisi aina ilmaista minkä tyyppisiä tietoja aineistossa on rajoitettu, miten rajoitus on ilmaistu aineistossa, mihin (säädökseen) rajoitus perustuu ja miten rajoitettuun osaan saa käyttöoikeuden.

Embargo aika on käytössä myös Lajitietokeskuksen aineistopolitiikassa. Erytistapauksen muodostavat pitkäaikaisia aikasarjoja muodostavat seuranta-aineistot, joiden embargoajaksi on määritelty neljä "uusinta" vuotta.

Lajien levinneisyyttä kuvaavia aineistoja jaetaan kansainvälisesti Dublin Coreen perustuvaa Darwin Core (DwC) –standardia⁴⁷ käyttäen. DwC mahdollistaa yksittäisen tiedon piilottaminen <information withheld>- "tagillä", jolloin tiedon olemassaolo ja rajoitusperuste tulee näkyville myös yksittäistä tietuetta haettaessa. Tämä mahdollistaa täydellisen aineiston käyttämisen eri sopimuksella tutkimuksessa, kunhan sovitaan siitä, miten rajoitus huomioidaan tutkimustulosten tai tutkimusaineiston edelleen julkaisemisessa.

Metatietojen lisäksi aineisto voi siis olla vain osittain julkinen, jolloin metatiedoista tulisi käydä ilmi mitä osia aineistoista on saatavissa erillisellä tutkimussopimuksella ja mitä ehtoja aineiston tai tulosten julkaisemiseen liittyy.

⁴⁷ Darwin Core <http://rs.tdwg.org/dwc/>

Oman ongelmakenttensä aineistojen avaamiselle tuovat kansainväliset arkistopalvelut ja data-portaalit. GBIF eli Global Biodiversity Information Facility pitää yllä dataportaalia, jonka kautta lajien levinneisyyttä kuvaavia tietoja voi julkaista globaalisti. Mekanismia käytetään menestyksekkäästi mm. siirtomaiden alueilta kerättyjen museonäytteiden tietojen luovuttamisessa takaisin myös alkuperämaan tutkimuksen käyttöön.

Esimerkkejä tutkimusaineistoista

Aalto yliopiston Brain & Mind Laboratory:

Elokuvan käyttö aivotutkimuksessa

Brain & Mind Laboratory (BML) on Aalto-yliopiston neurotieteen ja lääketieteellisen tekniikan laitoksen tutkimusryhmä, jossa tehdään tutkimusta aivo- ja hermomekanismien vaikutuksesta ihmisen mieleen. Tutkimuksessa karakterisoidaan käyttäytymistä, arvioidaan subjektiivisia kokemuksia, kerätään taustatietoja potilaista ja mitataan aivojen toimintaa joko toiminnallisella magneettikuvauksella (fMRI), magnetoenkefalografialla (MEG) tai aivosähkökäyrällä (EEG).

Aineistossa "Elokuvan käyttö aivotutkimuksessa" koehenkilöiden aivoja on kuvattu toiminnallisella magneettikuvauksella, samalla kun heille on näytetty valittua elokuvaa. Toiminnallisessa magneettikuvauksessa tutkitaan ihmisen aivoja luonnollisissa olosuhteissa eli esimerkiksi äänen tai liikkuvan kuvan ärsykkeiden vaikutuksessa. Tämä on erityisen hyödyllistä siksi, että tällöin mittauksia pystytään tarkasti analysoimaan ja vertaamaan ärsykkeeseen.

Lääketieteellisessä kokeessa tuotettua aineistoa, joiden tuottamiseen on saatu lupa eettiseltä toimikunnalta (Uudenmaan sairaanhoitopiirin Lasten ja nuorten sairauksien eettinen toimikunta). Aivojen kuvantamisaineistojen julkaisemiseen liittyvät juridiset vastuut ovat epäselvät. BML päättää aineiston jakamisesta eteenpäin tapauskohtaisesti, koska yhtenäistä käytäntöä jakamisesta ei Suomessa ole.

Aineistoon kuuluu myös muokattu versio Aki Kaurismäen elokuvasta "Tulitikkutehtaan tyttö". Elokuvaa sallitaan vain tutkimuskäyttöön, eli sen jakelu vapaasti ei ole sallittua.

Jyväskylä yliopiston Kiihdytinlaboratorio:

250-Nobeliumin hajoamisspektroskopia

Kiihdytinlaboratorio on kansallisesti ja kansainvälisesti merkittävä tutkimusinfrastruktuuri, jossa tehdään pääosin ydinfysiikan perustutkimusta. Kokeiden valmistelu ja suorittaminen on usein varsin arvokasta taloudellisesti ja datamäärät erittäin suuria, jopa kymmeniä teratavuja viikossa.

Aineistossa "*250-Nobeliumin hajoamisspektroskopia*" tutkitaan superraskaiden alkuaineiden radioaktiivisia hajoamistapahtumia ja se on mitattu käyttäen kiihdytinlaboratorion RITU-rekyyli-separaattoria ja GREAT-spektrometriä. Aineiston koko on 200 gigatavua ja mittaus on tehty maaliskuussa 2015.

Aineiston omistajuuden ja käyttöoikeuksien määrittely osoittautuivat vaikeaksi. Mittausaineistot on vapaasti käytössä mittauksiin osallistuneiden kesken. Kirjallisia sopimuksia aineiston omistajuudesta ei vaadita. Aineistoja ei ole avattu tai lisensoitu mittauksen ulkopuolisille tahoille.

Turun yliopiston Avaruustutkimuslaboratorio:

ERNE-instrumentin vuohavainnot 1996-2014 energioissa 1.6-200 MeV/n.

Aineisto on avointa ja vapaasti jaettavissa. Aineiston avoimuus on SOHO-observatorion laukaisusta vastanneiden ESA:n ja NASA:n vaatimus. Erityisen haasteellisia olivat julkaisujen käyttöoikeusasiat, koska julkaisujen liittäminen aineistoihin edellyttää kustantajien luvat, joiden saaminen osoittautui haastavaksi. Aineisto itsessään lisensoitiin Creative Commons-lisenssillä.

STRATAS

Interfacing structured and unstructured data in sociolinguistic research on language change. Suomen Akatemian tutkimushanke.

Hankkeessa tuotettavan kirjekorpuksen on tarkoitus kuvata mahdollisimman hyvin sosiaalista todellisuutta siten, että siinä olisi eri sosiaalisten ryhmien edustajia, koulutettua eliittiä, kouluttamattomia tai vähän koulua käyneitä, naisia ja miehiä, nuoria ja vanhoja, eri puolilta Suomea, kaupungista ja maaseudulta, siirtolaisia ja heidän sukulaisiaan jne. Parhaillaan kootaan kirjeitä eri suunnista, käydään läpi arkistoja ja yritetään saada aineistoa myös muilta tutkijoilta, jotka olisivat valmiita luovuttamaan omaa aineistoaan muiden käyttöön. Tässä tarkoituksessa lähestytään mm. 1800-luvun tutkijoiden verkostoa.

Kirjeiden on tarkoitus olla valokuvana, kielentutkimukseen käsiteltynä aineistona sekä sellaisena tavallisena tekstiaineistona, jota kuka tahansa voi lukea. Konsortiolla tulee todennäköisesti olemaan jokin alusta, jossa aineisto on saatavissa, ja luonnollisesti pohditaan myös mahdollisuutta luovuttaa tekstiaineisto myös Kielipankkiin. Tarkoitus olisi, että kirjekorpus olisi mahdollisimman avoin kaikkien alojen tutkijoille, ei vain kielitieteilijöille. Kirjeisiin merkitään esim. kohdat, joissa mainitaan henkilön- tai paikannimiä, mikä voi olla esim. kulttuurihistorialliseen tutkimukseen kiinnostavaa tietoa.

Haasteita tuovat mm. tekijänoikeuslaki ja henkilötietolaki. Aineiston pitäisi olla sellaista, että tekijän kuolemasta on kulunut 70 vuotta ja kirjeen kirjoittamisesta 125 vuotta, mikä ei uusimman aineiston osalta tietenkään ole aina tilanne (esim. vuonna 1900 syntynyt ihminen, joka kirjoittaa nuorena, ennen 1921). Ongelmana on myös se, että koska ideana on yhdistää teksti rakenteiseen tietoon, kirjekorpuksen rinnalle on koottava erillinen metadatatiedosto, joka sisältää tietoja kirjoittajista ja vastaanottajista. Näin luodaan myös eräänlainen henkilötietokanta, vaikka sen ei tietenkään ole tarkoitus olla julkinen, vaan tutkijoille luvanvarainen.⁴⁸

Pitkäaikaissaatavuus

Pitkäaikaissäilytyspalvelun (PAS) tarkoituksena ei ole pelkästään säilyttää bittejä, joista digitaaliset aineistot koostuvat, vaan turvata myös niiden ymmärrettävyyden säilyminen. Tämä mahdollistaa aineistojen joustavan hyödyntämisen hyvin erilaisiin tarkoituksiin myös tulevaisuudessa. Kirjastojen, arkistojen ja museoiden sähköisten aineistojen saatavuuden varmistamiseksi on jo vuonna 2008 aloitettu hanke Kansallisen digitaalisen kirjaston osana. KDK-PAS-palvelun tehtäviin kuuluu kehittää toiminnallisuuksia, jotka edistävät aineistojen käyttöä, ja seurata pitkäaikaissäilytyksen yleistä kehitystä. Pitkäaikaissäilytykseen liittyvien prosessien, menetelmien ja mallien kehitys tehdään läheisessä yhteistyössä mukana olevien arkistojen, kirjastojen ja museoiden kanssa. Yhteistyöllä varmistetaan työn tuloksien soveltuvuus kulttuuriperintöaineistojen säilyttämiseen ja erityisesti hyödyntämiseen.

KDK:n PAS-palvelun alustana toimii keskitetty pitkäaikaissäilytysratkaisu (PAS-ratkaisu), joka keskittyy erityisesti pitkäaikaissäilytyksen bittitason toiminnallisuuteen ja prosesseihin ja jonka avulla voidaan tuottaa PAS-palveluita eri tarpeisiin.

⁴⁸ Taru Nordlund, sähköposti, 6.4.2016.

KDK-PAS-palvelun ja PAS-ratkaisun toteutuksesta vastaa Tieteen tietotekniikan keskus CSC, ja niitä kehitetään tiiviissä yhteistyössä Avoimen tieteen ja tutkimuksen hankkeen kanssa.⁴⁹

Tutkimuksen materiaalien pitkäaikaisen säilymisen ja saatavuuden varmistaminen on myös yksi ATT-hankkeen tavoitteista. Hankkeen tiekartassa määritelty tehtäväksi ”tutkimuksen tulosten pitkäaikaissäilytyksen kehitys” vuosina 2015–2016. Tiekartalla on lähdetty siitä, että tutkimusaineistojen pitkäaikaissäilytys on toiminnassa vuonna 2017. ATT-hankkeelle tehdään vuosittaiset toimintasuunnitelmat. Tätä tavoitetta toteuttaa osaltaan tutkimuksen PAS-palvelu. Tutkimuksen PAS-palvelu tarjoaa keinot tutkimusdatan ja muiden tutkimuksen tuotosten säilytyskelpoiseksi kokonaisuudeksi saattamiseen, tallentamiseen sekä uudelleen hyödyntämiseen.

PAS-ratkaisu tuottaa säilymisen ydinpalvelun tutkimuksen PAS-palvelulle erikseen sovittavalla tavalla. Tämä seikka vaikuttaa tietenkin myös vaatimusmäärittelyyn, ja erityisesti Tutkimuksen PASin vaatimukset vaikuttavat PAS-ratkaisun kehitykseen.⁵⁰

Tutkimusaineistojen pitkäaikaissaatavuuspalvelun valmisteluvaiheessa huolehditaan, että aineistojen alkuperäiset bittijonot pystytään säilyttämään muuttumattomana ja voidaan toistaa nykyaikaisilla laitteilla. Bittien säilyttämisen nopealla käynnistämällä varmistetaan, että palvelua hyödyntävien organisaatioiden hallussa olevat digitaaliset aineistot pystytään luotettavasti säilyttämään siihen asti kunnes palvelu kokonaisuudessaan on tuotannossa. PAS-ratkaisun toteutuksen toisessa vaiheessa huolehditaan aineistojen ymmärrettävyyden säilymisestä, niiden informaatioisällön tulkittavuudesta ja aineistojen hyödynnettävyydestä myös tulevien vuosikymmenien ohjelmistoilla. Ymmärrettävyyden säilyttämisen toteuttaminen aloitetaan vasta sitten kun bittien säilyttäminen on saatu tuotantokäyttöön.⁵¹

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.⁵²

Yleisenä metatietojen poiminnan rajapintamallina sovelletaan KDK-hankkeessa ja laajemminkin kirjasto- ja arkistoaloilla OAI-PMH -protokollaa.⁵³ 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

⁴⁹ Pitkäaikaissäilytys <http://www.kdk.fi/index.php/fi/pitkaaikaissailytys>

⁵⁰ Tutkimuksen pitkäaikaissäilytys, keskeneräinen vaatimusmäärittely 0.2

⁵¹ Pitkäaikaissäilytyksen toteuttaminen <http://avointiede.fi/toteuttaminen>

⁵² Ks esimerkiksi <http://iow.csc.fi/#/>

⁵³ The Open Archives Initiative Protocol for Metadata Harvesting <http://www.openarchives.org/OAI/openarchivesprotocol.html>

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 metadatasuosituksia, jonka odotetaan valmistuvan vuoden 2017 alkupuolella.⁵⁴ Suosituksessa pyritään luomaan yhtenäiset, mahdollisimman yksinkertaiset käytännöt myös käyttöoikeuksien ilmaisemiselle julkaisuarkistojen sisäisessä metadatformaatissa. Tavoitteena on, että julkaisuarkistoista OAI-PMH-rajapinnan ulospäin välitettäviin metadatformaatteihin voidaan tallennettujen tietojen pohjalta generoida tarpeen mukaan eri tavoilla muotoiltuja tietoja.

Finnaan⁵⁵ 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 pitkäaikaissäilytykseen 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.

⁵⁴ Ks. TAJUA-yhteistyöverkosto <https://www.kiwi.fi/x/f4OFAw>

⁵⁵ Aineistojen haravointi Finnaan Finna-wikissä <https://www.kiwi.fi/x/PA0XAw>

4. 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 käyttöoikeuksia 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.

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 geodalalle ja tilastoille. Myös ATT-hankkeessa on huomioitu DCAT-yhteensopivuus. DCATiin liittyy myös W3C:n suosittelema sanasto.⁵⁶

DCAT on siis keskeinen, mutta myös vasta kehittyvä ohje. Käyttöoikeuksien metatietojen tallennusta varten standardissa käytetään Dublin Core

⁵⁶ DCat vocabulay <https://www.w3.org/TR/vocab-dcat/>

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 Vocabularyssa⁵⁷ sekä Open Digital Rights Languageissa (ORDL).⁵⁸ Myös lisenssien luokittelua ollaan tekemässä.⁵⁹ Aikaisemmin luotu eurooppalaisten info:eu-repo -yhteentoimivuustyö on loppunut, mutta EU:ssä käytetään edelleen sen luomia saatavuusluokituksia tai niiden muunnelmia.⁶⁰

Europeana

Europeanan ja DPLA:n Rightsstatements.org -yhteistyö on ehdottanut seuraavia luokkia tekijänoikeuksiin liittyvien oikeuksien suhteen (White paper updated in January 2016).⁶¹ Marraskuussa 2016 Europeana otti käyttöön näistä kuusi eri arvoa (keltaisella pohjalla):⁶²

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.
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.

⁵⁷ Open Data Institute. Open Data Rights Statement Vocabulary. <http://schema.theodi.org/odrs/>

⁵⁸ 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>.

⁵⁹ https://joinup.ec.europa.eu/asset/dcat_application_profile/description ; <https://www.w3.org/TR/vocab-dcat/>

⁶⁰ Info-eu-repo, Wiki <https://wiki.surfnet.nl/display/standards/info-eu-repo#info-eu-repo-AccessRights>

⁶¹ Rightsstatements.org White Paper http://rightsstatements.org/files/160208recommendations_for_standardized_international_rights_statements_v1.1.pdf

⁶² Europeana Factsheet pro.europeana.eu/files/Europeana_Professional/IPR/rightsstatement.org/europeanafactsheet.pdf

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,⁶³ kun ne kerran on saanut. Sen tähden myös erillinen tieto pääsystä on tarpeen.

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.⁶⁴ 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 kehittyy 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-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

⁶³ 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/>

⁶⁴ Legal Interoperability Interest Group RDA

<https://rd-alliance.org/groups/rdacodata-legal-interoperability-ig.html>

Elements (ESE). EDM on yhteensopiva ESE:n kanssa ja Europeana tukee edelleen myös ESE-muotoista dataa.⁶⁵

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.⁶⁶

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.⁶⁷

Yhdysvalloissa ANSI/NISO (American National Standards Institute / National Information Standards Organization) on tuottanut Access License and Indicators -suosituksen⁶⁸. 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.
[---]

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.

Ohjeistukseen sisältyy myös esimerkkejä XML-skeemoista joilla suosituksen mukaiset metatiedot saadaan koneluettaviksi.

⁶⁵ Lisätietoa <http://pro.europeana.eu/page/edm-documentation>

⁶⁶ Formula Kansallisen digitaalisen kirjaston sivuilla
<http://www.kdk.fi/index.php/fi/europeana-ja-muut-hankkeet/formula>

⁶⁷ DataCite Metadata Kernel
http://schema.datacite.org/meta/kernel-3/doc/DataCite-MetadataKernel_v3.1.pdf

⁶⁸ 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

5. Suositukset oikeuksien hallintaan liittyvien metatietojen luomiseen ja käyttöön

1. Avoimuus on aina lähtökohta. Ellei aineiston tai metadatan käyttöä tai saatavuutta ole rajoitettava lain, sopimuksen tai eettisen syyn perusteella sen tulisi olla avointa. Avoimuuteen ja kestäväään käyttöoikeuksien hallintaan on pyrittävä aineiston koko elinkaaren ajan.
2. Käyttö- ja saatavuus- eli pääsyrajoitukset on erotettava toisistaan. Käytä olemassa olevia, soveltuvia standardeja ja kuvailuohjeita.
3. Kaikki rajoitukset tulisi ilmoittaa metatiedoissa selvästi, mielellään sekä ihmis- että koneluettavalla ja -ymmärrettävällä tavalla.
4. Käyttörajoitusten tarkan kohteen tulisi 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 ihmis- että koneluettavalla tavalla. (Ehdotus liitteessä 3)
6. Tekijänoikeuksiin perustuvat käyttörajoitukset tulisi, jos mahdollista, kertoa Creative Commons-lisenssien avulla. Jos ne eivät sovellu, voidaan käyttää muita laajalti käytettyjä ratkaisuja, kuten Europeanan ja DPLA:n rightsstatements -suosituksia (http://rightsstatements.org/files/151002recommendations_for_standardized_international_rights_statements.pdf). Metatietoihin tulisi tallentaa lisenssin nimi tai sen lyhennelmä sekä - jos mahdollista - linkki lisenssin (suomenkieliseen) tiivistelmään tai, ellei sitä ole käytettävissä, kokotekstiin. (Ehdotus liitteessä 3)
7. Mikäli käytössä on organisaatio- tai aineistokohtaisia käyttörajoituksia niitä tulisi käsitellä seuraavalla tavalla:
 - Tarjoa aina ihmisen luettavaa versiota lisenssistä [uri:n takana]
 - Kuvaile mitä käyttäjä saa tehdä aineistolle
 - Kuvaile mitä ehtoja tähän liittyy
 - Jos mahdollista tee vain yksi versio lisenssin sisällöstä
 - Linkitä ensin helppolukuiselle sivulle ja sieltä koko lisenssiin/lakiin
 - Määrittele käsitteet, ajattele luettavuutta
 - Tarjoa sekä html- että pdf-versio
8. Rajoituksen perusteen tyyppi tulee ilmaista metatiedoissa. (Ehdotus liitteessä 3)
9. Oikeuksien haltijat on ilmoitettava mahdollisimman yksiselitteisesti.
10. Aikarajoitukset ja päiväykset on huomioitava metatiedoissa ja ne on kerrottava yksiselitteisesti ja standardien mukaisesti. Esimerkiksi päivämäärät on kuvattava ISO 8601 -standardin esittämällä tavalla (VVVV-KK-PP).

6. Etenemisehdotus

Tavoitteet
1. Oikeuksien hallintaa koskevien metatietojen suositusten ja aineiston kokoaminen
2. Aineistoihin liittyviä vastuuta ja periaatteita on edelleen selkeytettävä <ul style="list-style-type: none"> • Aineistojen elinkaaren suunnittelussa on otettava huomioon prosessin eri vaiheet. Olisi hyvä vaikuttaa käytäntöihin siten, että eri vastuiden ja oikeuksien määrittely tehtäisiin huolellisesti. • Aineistojen omistus tulisi olla mieluummin organisaatiolla kuin yksittäisellä henkilöllä tai ryhmällä kuten tutkijalla tai tutkimusryhmällä • Luvanvaraisten aineistojen kohdalla luvan antajan (organisaatio) mielekäs kertominen metadatatassa vaatii poliittis-hallinnollisia rakenteita, joita pitää selvittää ja kehittää varsinkin tutkimussektorilla.
3. Suositusten jatkotyöstäminen ja tarkentaminen kansallisella tasolla <ul style="list-style-type: none"> • Yleisesti käytetyille metatietoformaateille tulisi laatia ohjeet, jotka kuvaavat miten käyttöoikeustiedot tallennetaan. Ohjetta voidaan käyttää hyväksi sekä kuvailutyön ohjauksessa että sovelluskehityksessä. Näitä ohjeita tulisi tehdä eri sektoreilla, mutta ne pitäisi myös koota. Ohjeissa pyritään seuraamaan alla olevia suosituksia niin pitkälle kuin mahdollista. Käytännössä formaateissa ja varsinkaan järjestelmissä niitä ei vielä ole täysin mahdollista toteuttaa. • Ohjeet laaditaan sektori- ja formaattikohtaisesti kattavin esimerkein. Tietojärjestelmiä valittaessa ja kehitettäessä oikeuksiin liittyvät metatiedot on huomioitava. • Tutkimusaineistojen kuvailuformaateille laaditaan kansallinen ohjeistus jossa otetaan oikeuksien hallinnan metatietojen tallennus huomioon • Myös metatiedon tuottajille tulisi tuottaa ohjeita eri tilanteisiin (kuvailutyö, koulutus, opastus jne). Kaikki (linkit) voidaan koota yhteiseen verkkoresurssiin. Myös tutkijoille suositellaan (esim. ATTN kautta) että kun valitaan lisenssi, valitaan mieluummin CC-lisenssi kuin ns. kustantajan oma lisenssi, jos mahdollista. Ohjeistuksen kehittämistä jatketaan yhteistyössä, sillä tilanne kehittyy jatkuvasti. Työtä jatketaan roolien ja oikeuksien määrittelyllä, kun tilanne on niiden kohdalla selkeytynyt.
4. Suositusten vieminen kansainvälisiin formaatteihin, sekä kansainvälisen kehityksen seuranta <ul style="list-style-type: none"> • Kansainvälistä kehitystä on seurattava, sekä pohjoismaista ja suomalaista näkökulmaa on tuotava aktiivisesti esille, jotta kehittyvät kansainväliset standardit sopivat myös meille.
5. Oikeuksia käsittelevät sanastot Fintoon <ul style="list-style-type: none"> • Keskeisistä lisensseistä, tekijänoikeudellisista statuksista ja saatavuusrajoituksesta sekä rajoitusten perusteiden tyypeistä (liite) tehdään sanasto heti kun ne on hyväksytyt. Sanasto viedään Fintoon. Tämän olisi hyvä tapahtua mahdollisimman ripeästi. • Edistetään pysyvien tunnisteiden käyttöä ja laadukasta hallinnointia. • Käännetään tarvittavat lisenssitekstit
6. Käsitteet osaksi ATT-tietomallia ja käytäntöä tutkimuspuolella <ul style="list-style-type: none"> • Sanastotyö luovutetaan ATT:n ja TUHA:n tietomallityöhön semanttisen yhteentoimivuuden varmistamiseksi. • Kehitetään Etsintä käyttäjälähtöisesti tietomallityön tukemiseksi

7. Lainsäädännön muutoksia pitäisi edistää

- Ainakin henkilötietoja ja tekijänoikeuksiin kuuluvia asioita, joista tarkemmin tekstissä, jotka vaikeuttavat organisaatioiden ja tutkijoiden työtä olisi tärkeä selvittää edelleen. Lainsäädäntöön tarvittavia muutoksia tulisi edistää. Olisi tärkeää turvata tieteellisten aineistojen, mukaan lukien kulttuuriperintöaineistot, tarkoituksenmukainen säilytys ja saatavuus.

8. Identiteetinhallinnan kehittäminen

- Olisi syytä kehittää kansallista sähköistä identiteetin hallintaa. Kattava identiteetinhallinta on edellytys manuaalisen työn vähentämiselle ja palveluiden integroinnille. Myös tutkijoiden kansainvälinen liikkuvuus on otettava huomioon.
- Myös vahvan tunnistautumisen tulisi olla mahdollista käyttörajoitettujen aineistojen ja metatietojen saatavuuden parantamiseksi.

-