

2015

Tiedon avaaminen Museovirastossa

OHJEITA JA TIETOA DATAN AVAAMISEN TUEKSI, 22.5.2015
PIA LONARDI, MIIKKA HAIMILA, TAPANI SAINIO, ANNIKA TUOMALA

Sisällysluettelo

Johdanto	4
Avoin tieto.....	5
Alan verkostoja Suomessa	6
Open Knowledge Finland.....	6
AvoinGLAM.....	7
Wikimedia Suomi	7
Avoin hallinto -hanke	8
Creative Commons -lisenssit.....	8
Tiedon avaaminen Museovirastossa.....	12
Museoviraston hallinnoimat tietoaineistot.....	13
Datan avaajan tarkistuslista	13
Datan avaamiseen liittyvä lainsäädäntö.....	14
Aineiston saattaminen julkaistavaan muotoon.....	17
Tietotuoteseloste ja sen laadinta.....	18
Tietoaineistojen ylläpito ja päivittäminen	18
Tietoaineistojen käytön edistäminen	19
Case: Museotilasto	19
Hankkeen taustat Museovirastossa ja tiedon avaamisen tavoitteet	20
Tilastodatan avaamiseen liittyvä lainsäädäntö.....	21
Avaamisen vaiheet.....	22
Aineiston valinta ja sisäisen työryhmän muodostaminen.....	22
Julkaistavien tietojen laajuus	22
Museoille suunnattu kommentointikierrös.....	23
Tietotuoteseloste ja sen sisältö	23

Suostumus tietojen julkaisuun.....	24
Käytännön toteutus.....	24
Lopuksi.....	25
Kirjallisuutta ja tietolähteitä.....	26

Tämä opas on lisensoitu Creative Commons Nimeä 4.0 Kansainvälinen (CC BY 4.0) lisenssillä. Oppaan uudelleenkäytön yhteydessä tulee mainita sekä kirjoittajat (Pia Lonardi, Miikka Haimila, Tapani Sainio ja Annika Tuomala) että organisaatio, jonka piirissä opas on tuotettu (Museovirasto). Lisenssiin voi tutustua tarkemmin: <https://creativecommons.org/licenses/by/4.0/deed.fi>

Johdanto

Julkisen hallinnon tuottamat tietovarannot sekä erilaisten kulttuuriperintöorganisaatioiden hallussa olevat sisällöt ovat olleet ajankohtaisen keskustelun aiheena Suomessa ja Euroopan Unionissa jo useamman vuoden ajan. Yhteiskunnan digitalisoituessa digitaalisten tietovarantojen ja palveluiden merkitys korostuu sekä julkisen että yksityisen sektorin näkökulmasta. Parhaimmillaan avointen sisältöjen ja avoimen toimintakulttuurin kautta voidaan edistää demokratiaa, kansalaistoimintaa ja liike-elämää. Näin aineistojen avaaminen hyödyttää usein myös tietovarantoja hallinnoivia organisaatioita itseään.

Museovirastoa keskustelu koskee niin valtion virastona kuin kulttuuriperintöorganisaationa. Tietoaineistoja voidaan jaotella eri tavoin. Tiedon avaamisen toimijoiden ja verkostojen näkökulmasta Museovirasto tuottaa sekä viranomaistietoa että kirjasto-, arkisto ja museotoimintaan liittyvää kulttuuridataa. Erityyppisiä tietovarantoja kannattaa lähestyä hieman eri tavoin ja yhteistyössä erilaisten toimijoiden tai verkostojen kanssa. Tietoa voidaan julkaista verkossa avoimen tiedon lisäksi myös rajatumilla käyttöoikeuksilla. Tällöinkin on tärkeää merkitä aineistoon mahdollisesti liittyvät käyttörajoitukset selkeästi.

Datan avaamisessa kannattaa lähteä liikkeelle pienistä ja mahdollisimman yksinkertaisista kokonaisuuksista, minkä vuoksi Museovirastossa kohteeksi valikoitui Museotilaston kautta tuotettu aineisto (museotilasto.fi). Aiemmin on Museoviraston aineistoista avattu muun muassa paikkatietoaineistoa, Europeanaan ja Finnaan välitettyä kuvailutietoa sekä kuva-aineistoa eri hankkeiden yhteydessä.

Tämä opas perustuu tilastodatan avaamisen yhteydessä kerättyyn tietoon ja projektin aikana opittuihin asioihin, mistä syystä näkökulma ja käytännön esimerkit keskittyvät Museotilastoon. Museoviraston tehtävänä on suomalaista museoalaa koskevan toimialatilaston kerääminen. Tilastoa kerätään verkkopalvelun avulla vuosittain ammatillisesti hoidetuilta museoilta. Museotilastoa on kerätty sähköisesti vuodesta 2007 lähtien. Sähköisen palvelun vahvuutena voidaan pitää tiedon keräämisen tehokkuutta, mutta ennen kaikkea tiedon saavutettavuuden paranemista. Tilastotiedon lisensointi eli käyttöluvan määrittely sekä tiedon julkaiseminen avoimena datana koettiin luonnollisena keinona saavutettavuuden parantamiseksi.

Tietoaineistojen lisensoinnin avulla on mahdollista selvittää tiedon käyttöoikeuksia, mikä samalla helpottaa tietojen käyttöä. Julkisen hallinnon tietohallinnon neuvottelukunnan (JUHTA) sekä sen alaisen

[JHS-jaoston suosituksen](#) mukaisesti Museotilastossa kerättyyn tietoon päätettiin myöntää käyttöoikeudet avoimella käyttöluvalla [Creative Commons Nimeä 4.0 Kansainvälinen \(CC BY 4.0\)](#).

Museotilastossa kerätyn tiedon osalta se tarkoittaa, että tiedon käyttöä, jakelua ja muokkausta ei rajoiteta millään tavalla, mutta käyttäjän tulee viitata alkuperäiseen tietolähteeseen, eli Museoviraston tuottamaan Museotilastoon ja merkittävä, mikäli aineistoon on tehty muutoksia. Näin turvataan loppukäyttäjän mahdollisuudet alkuperäisen tiedon löytämiseen sekä jaetun tiedon oikeellisuuden tarkistamiseen.

Tämä on ensimmäinen versio Tiedon avaaminen Museovirastossa –oppaasta. Kun kokemus datan avaamisesta Museovirastossa karttuu, on oppaan täydentäminen ja muokkaaminen erittäin tervetullutta.

Avoim data

Avoim data ei ole sama asia kuin julkinen tieto. Julkiseen tietoon kaikilla on pääsy, eli ihmiset pääsevät **lukemaan tietoja** esim. verkkosivuilta tai kaupungin kirjaamosta. Avoin julkinen tieto eli avoin data puolestaan tarkoittaa sitä, että kansalaiset ja yritykset voivat **käyttää tietoja omiin tarkoituksiinsa** tasavertaisesti julkisen hallinnon kanssa.¹

Open Knowledge Foundation on [määritellyt avoimuuden seuraavasti](#):

- Aineiston tulee olla kokonaisuudessaan saavutettavissa ja ladattavissa julkisessa tietoverkossa.
- Tiedon on oltava uudelleenjaettavissa ilman käyttöehtorajoituksia. Näin mahdollistetaan aineistojen nopea hyödyntäminen ja tehokas yhdistely.
- Tiedon on oltava uudelleenkäytettävissä. Näin sallitaan aineistojen esteetön ja innovatiivinen käyttö edistykseen ja yllättäviinkin tarkoituksiin.
- Aineiston on oltava vapaa teknisistä rajoitteista niin, ettei yllämainittujen kohtien mukaiselle toiminnalle ole teknisiä esteitä.
- Aineiston on oltava vapaa sosiaalisista ja organisatorisista rajoitteista, niin ettei henkilön työ, sijainti, asuinpaikka, organisaatiomalli (kuten kaupallinen tai ei-kaupallinen organisaatio), uskonto, poliittinen suuntautuneisuus tai etnisuus rajoita pääsyä tietoon.²

Tietojen avaajille on erilaisia työkaluja, jotka auttavat hahmottamaan, mitä avoimuus käytännössä vaatii.

Tim Berners-Lee on ehdottanut [viiden tähden avoimen datan mallia](#), jossa määritellään datan avoimuus eri tasoilla.³ Tässä mallissa avoimuus keskittyy ennen kaikkea siihen, missä tiedostomuodossa ja millä

¹ "Mitä on avoin data?", Helsinki Region Infoshare. <http://www.hri.fi/fi/mita-on-avoin-data/> 25.2.2015.

² "Visio ja arvot", Open Knowledge Finland. <http://fi.okfn.org/about/visiojaarvot/> 25.2.2015.

³ "5 * Open data". <http://5stardata.info/> 25.2.2015.

standardeilla avoin data julkaistaan. Taulukossa⁴ kuvataan eritasoista avointa dataa ja annetaan esimerkkejä eri tiedostomuodoista ja standardeista, joiden puitteissa dataa voi julkaista.

-	Aineisto voi olla saatavilla verkossa, mutta sitä ei ole julkaistu avoimella lisenssillä	
*	Aineisto on saatavilla avoimella lisenssillä missä tahansa tiedostomuodossa	Esimerkiksi pdf-dokumentti tai kuvakaappaus aineistosta
**	Aineisto on lisäksi saatavilla rakenteisessa ja koneluettavassa tiedostomuodossa	Esimerkiksi Excel-tiedosto
***	Aineisto on lisäksi julkaistu avoimena tiedostomuotona	Esimerkiksi CSV-tiedosto (comma-separated values). Tiedostomuotoon tallennetaan yksinkertaista taulukkomuotoista tietoa tekstitiedostona.
****	Aineistolla on lisäksi URI (Uniform Resource Identifier), josta ilmenee tiedon paikka tai yksikäsitteinen nimi. Identifiointi mahdollistaa aineiston linkittämisen toisiin lähteisiin.	Esimerkiksi RDFa-tiedosto (Resource Description Framework in Attributes), jossa on URIa.
*****	Aineisto on lisäksi linkitetty muihin aineistoihin, ja näin ollen toisten lähteiden löydettävissä ja kontekstualisoitavissa	Esimerkiksi RDFa-tiedosto, jossa on URIa ja semanttista tietoa (esim. Wikipedian linkkien kautta)

Alan verkostoja Suomessa

Suomessa toimii tällä hetkellä useita tiedon avaamiseen ja avoimen tiedon hyödyntämiseen keskittyviä verkostoja. Museoviraston kannalta keskeisiä toimijoita ovat ainakin [Open Knowledge Finland](#), [AvoinGLAM](#), [Wikimedia Suomi](#) ja [Avoin hallinto-hanke](#). Osallistumalla erilaisten verkostojen ja työryhmien toimintaan on mahdollista löytää yhteistyökumppaneita tietojen avaamisen ja hyödyntämisen tueksi.

Open Knowledge Finland

[Open Knowledge](#) on kansainvälinen avoimen tiedon tuotantoa ja käyttöä edistävä organisaatio. Suomen paikallisorganisaation, [Open Knowledge Finlandin](#) tavoitteena on yhdistää eri avoimen tiedon toimijoita (avoin lähdekoodi, avoimet lisenssit, avoin tiede, avoin sisältö jne.) ja löytää sekä tuottaa avoimen tiedon avaamiseen ja hyödyntämiseen tärkeitä projekteja Suomessa.⁵

Toiminta on jalkautettu erilaisiin työryhmiin: Open Government Data, Open GLAM (AvoinGLAM), Open Science, Open Sustainability, Open Education, Open Democracy, Open Licences ja My Data.

⁴ Marttila & Sillanpää, 2014. <http://avoinglam.fi/opas.pdf>

⁵ Finnish Open Data Ecosystem -Facebook-ryhmän kuvaus. <https://www.facebook.com/groups/fo.kfn/> 25.2.2015.

Open Knowledge verkosto on avannut [Datan avaamisen verkkokurssin](#). Kurssi sisältää paljon tietoa avoimesta datasta, datan avaamisesta ja aiheeseen liittyvästä lainsäädännöstä. Lisäksi kurssi sisältää harjoitustehtäviä.

AvoingLAM

[AvoingLAM](#) on Open Knowledge Finlandin alatyöryhmä ja osa kansainvälistä [Open GLAM](#) -verkostoa (GLAM = Galleries, Libraries, Archives, Museums). AvoingLAM keskittyy erityisesti kulttuuriorganisaatioiden piirissä tuotetun kulttuuridatan uudelleenkäytön edistämiseen. Työryhmän toimintaan on kaikilla avoin pääsy ja siihen liittyvää ajankohtaista keskustelua voi seurata esimerkiksi Facebookissa [AvoingLAM-ryhmässä](#).

*Verkosto on julkaissut erinomaisen "[Kohti avointa kulttuuria](#)" -oppaan, jossa kerrotaan muun muassa kulttuuriaineistojen avaamisesta ja avointen kulttuuriaineistojen hyödyntämisestä sekä avoimesta toimintakulttuurista.⁶ **Opas on kompakti ja käytännönläheinen esitys aiheesta ja siihen kannattaa tutustua ennen työhön tarttumista.***

Museoviraston kannalta AvoingLAMin kanssa tehtävä yhteistyö voisi olla kannattavaa erityisesti avattujen **kulttuuriaineistojen** hyödyntämisen edistämiseksi.

Wikimedia Suomi

[Wikimedia Suomi ry](#) tukee [Wikipedian](#) sekä muiden Wikimedia Foundationin ylläpitämien palveluiden tuntemusta ja käyttöä sekä vapaan sisällön luontia ja levittämistä Suomessa.

Wikipedia on lähes jokaisen verkosta tietoa hakevan ensisijainen tiedonlähde. Tämän vuoksi on myös tietoa tuottavien organisaatioiden ja tiedon jakamisen kannalta hyödyllistä, että Wikipediasta löytyvä tieto on ajantasaista ja mahdollisimman luotettavaa.

Asiantuntijaorganisaatioissa, etenkin Museoviraston museo- ja arkistopuolella, yhteistyö Wikimedia-verkoston kanssa voi olla monella tapaa hyödyllistä. Esimerkiksi nykytaiteen museo Kiasma on toteuttanut

⁶ "Kohti avointa kulttuuria -opas ilmestynyt", AvoingLAM blogi. <http://avoinglam.fi/?p=928> 25.2.2015.

mm. [Wikimaratonin](#), jonka aikana tuotettiin Wikipediaan tietoa nykyaikaisesta ja -taiteilijoista Kiasman laajaa kirjastoa hyödyntäen.

Avoim hallinto -hanke

Suomi hyväksyttiin vuonna 2013 mukaan kansainväliseen Avoimen hallinnon kumppanuusohjelmaan ([Open Government Partnership](#)). Samalla Suomi sitoutui [OGP:n julistukseen](#).

[Avoim hallinto -hanke](#) on Valtiovarainministeriön vetämä julkisen hallinnon avoimuuteen ja tiedon avaamisen edistämiseen tähtäävä kumppanuusohjelma, joka tuo yhteen eri ministeriöt, virastot ja kunnat. Hankkeen piirissä valmistellaan tällä hetkellä Suomen toista Avoimen hallinnon toimintasuunnitelmaa.

Museovirasto on mukana hankkeessa mahdollisuuksiensa mukaisesti. Hankkeen puitteissa tuotettuja työkaluja voi käyttää virkamiestoiminnan avoimuuden edistämiseksi ja verkoston tukea julkishallinnollisen datan avaamisessa.

Creative Commons -lisenssit

Tietoaineistoista tulee avointa dataa, kun niihin liitetään avoin lisenssi eli käyttöluva. Creative Commons lisenssien käyttö ja tietojen lisensointi ei vaadi erillisiä lupia tai rekisteröitymistä.⁷ Lisenssejä käytettäessä tulee kuitenkin varmistaa, että lisensoijalla on oikeudet tiedon käyttöoikeuden määrittelyyn ja datan avaamiseen tai tieto/sisältö on vapaata siten, että sitä ei enää esimerkiksi ikänsä tai luonteensa puolesta koske lainsäädännön rajoitukset, kuten tekijänoikeudet tai henkilötietojen suoja. Julkaistavan ja lisensoitavan tiedon tulee myös aina olla julkista, eikä se saa sisältää esimerkiksi salassa pidettävää tietoa.

[Käyttöehdot tulee tuoda selkeästi esiin julkaistun tiedon yhteydessä, jolloin tiedon mahdollinen käyttäjä tietää oikeutensa tiedon jatkokäytön suhteen.](#)

Tietojen lisensoinnin avulla voidaan selventää julkaistujen tietoaineistojen käyttöoikeuksia, mikä samalla helpottaa tietojen uudelleenkäyttöä ja niihin viittaamista. Julkisen hallinnon tietohallinnon neuvottelukunnan (JUHTA) sekä sen alaisen [JHS-jaoston suosituksen](#) mukaisesti julkishallinnollisen datan avaamisessa tulee mahdollisuuksien mukaan käyttää *Creative Commons Nimeä 4.0 Kansainvälinen* ([CC BY 4.0](#)) lisenssiä.

⁷ "Lisenssien käyttöohje", creativecommons.fi -blogi. <http://creativecommons.fi/lisenssit/lisenssien-kayttoohje/> 25.2.2015 & "Näin julkaiset avointa dataa", creativecommons.fi -blogi. <http://creativecommons.fi/2015/02/nain-julkaiset-avointa-dataa/> 25.2.2015.

*Creative Commons lisenssejä on kaiken kaikkiaan kuusi. Näistä avoimiksi lisensseiksi hyväksytään **CC BY** ja **CC BY-SA**.*

Nimeä (CC BY)

Tämä lisenssi sallii muiden levittää, remiksata ja tuunata teosta tai dataa sekä luoda sen pohjalta uusia teoksia, myös kaupallisessa tarkoituksessa, kunhan tiedon tuottaja tai teoksen tekijä mainitaan alkuperäisenä tekijänä. Tämä on lisensseistä kaikkein joustavin. Sitä suositellaan tilanteisiin, joissa teoksen käyttö ja leviäminen halutaan maksimoida.⁸

Nimeä-Jaa Samoin (CC BY-SA)

Tämä lisenssi sallii muiden levittää, remiksata ja tuunata teosta tai dataa sekä luoda sen pohjalta uusia teoksia, myös kaupallisessa tarkoituksessa, kunhan tiedon tuottaja tai teoksen tekijä mainitaan alkuperäisenä tekijänä ja uudet tuotokset lisensoidaan samoin ehdoin. Tätä lisenssiä verrataan usein "copyleft"-tyyppisiin vapaisiin ja avoimen lähdekoodin ohjelmistolisensseihin. Kaikki teokseen perustuvat uudet teokset pysyvät saman lisenssin alla, joten myös kaikkia johdannaisteoksia voidaan käyttää kaupallisesti. Wikipedia käyttää tätä lisenssiä, ja sitä suositellaan aineistoille, jotka voivat hyötyä Wikipedian tai vastaavasti lisensoitujen hankkeiden aineistojen sisällyttämisestä.⁹

Nimeä-Ei Muutoksia (CC BY-ND)

Tämä lisenssi sallii sekä kaupallisen että epäkaupallisen levittämisen, kunhan teosta levitetään kokonaisena ja muuttamattomana ja tiedon tuottaja tai teoksen tekijä mainitaan tekijänä.¹⁰

Nimeä-Ei Kaupallinen (CC BY-NC)

Tämä lisenssi sallii muiden levittää, remiksata ja tuunata teosta sekä luoda sen pohjalta uusia teoksia, mutta tämän kaiken kuitenkin vain epäkaupallisessa tarkoituksessa. Vaikka uusienkin teosten tulee olla epäkaupallisia ja teoksen tekijä on mainittava alkuperäisenä tekijänä, tällaisia johdannaisteoksia ei tarvitse lisensoida samoilla ehdoilla.¹¹

Nimeä-Ei Kaupallinen-Jaa Samoin (CC BY-NC-SA)

⁸ "Tietoa lisensseistä", CreativeCommons. <http://creativecommons.org/licenses/> 25.2.2015.

⁹ "Tietoa lisensseistä", CreativeCommons. <http://creativecommons.org/licenses/> 25.2.2015.

¹⁰ Ibid.

¹¹ Ibid.

Tämä lisenssi sallii muiden levittää, remiksata ja tuunata teosta sekä luoda sen pohjalta uusia teoksia, tämän kaiken kuitenkin vain epäkaupallisessa tarkoituksessa, kunhan teoksen tekijä mainitaan alkuperäisenä tekijänä ja uudet tuotokset lisensoidaan samoin ehdoin.¹²

Nimeä-Ei Kaupallinen-Ei Muutoksia (CC BY-NC-ND)

Tämä lisenssi on kaikkein rajoittavin kuudesta päälisenssistä ja sallii teoksen jakamisen muille, kunhan teoksen tekijä mainitaan tekijänä. Teosta ei tällöin kuitenkaan saa muuttaa millään tavoin, eikä sitä saa hyödyntää kaupallisesti.¹³

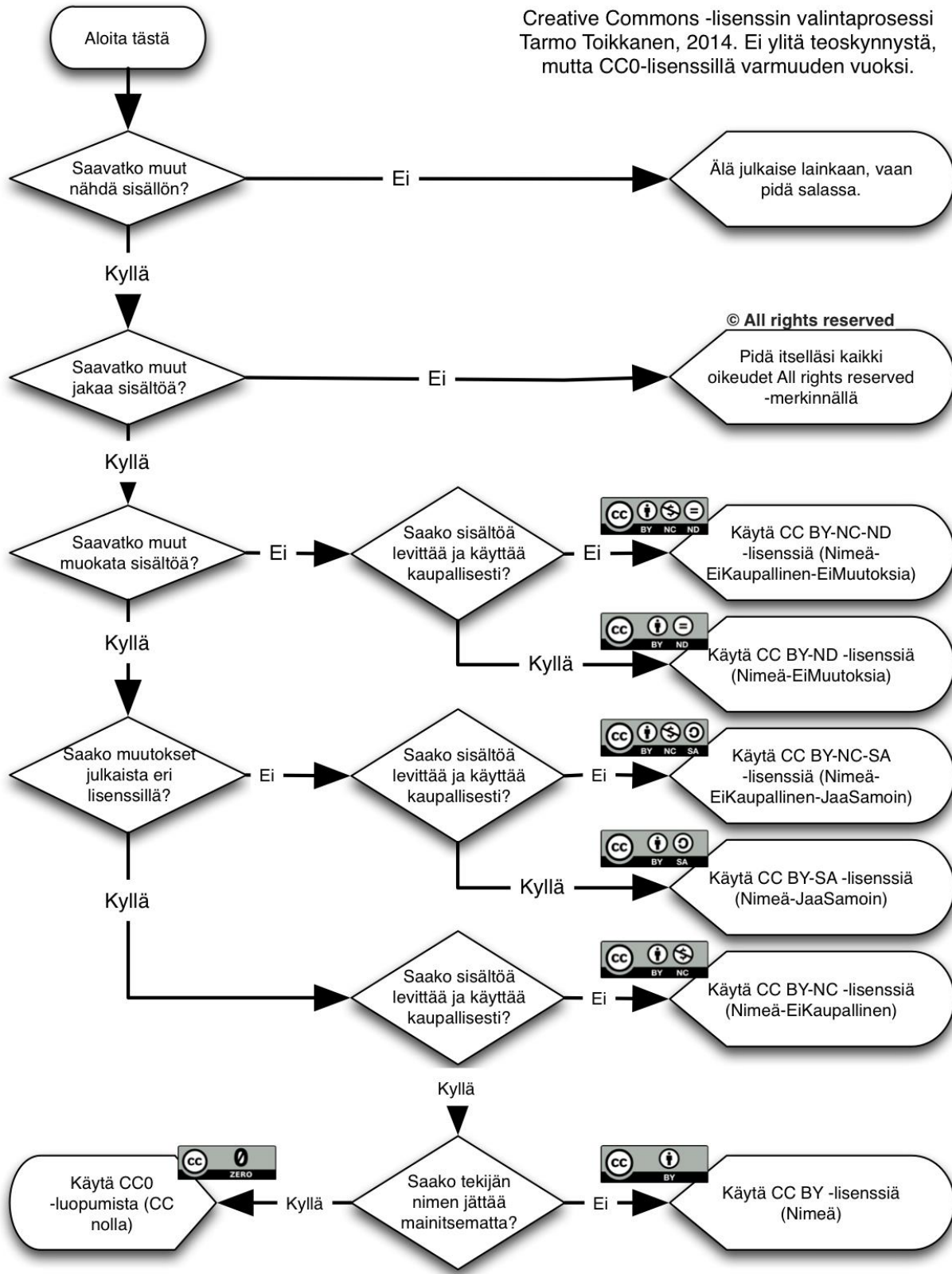
Lisenssin käyttöön liittyen tulee tehdä valintoja, jotka on kuvattu seuraavassa kuvassa:¹⁴

¹² Ibid.

¹³ Ibid.

¹⁴ "Valitse lisenssi", creativecommons.fi -blogi. <http://creativecommons.fi/lisenssit/valitse-lisenssi/> 25.2.2015.

Creative Commons -lisenssin valintaprosessi
Tarmo Toikkanen, 2014. Ei ylitä teoskynnystä,
mutta CC0-lisenssillä varmuuden vuoksi.



Tiedon avaaminen Museovirastossa

Datan avaaminen alkaa aina organisaation hallinnoimien tietovarantojen arvioinnilla. Kokonaiskuva Museoviraston hallinnoimista tietovarannoista on tärkeä, mutta kaikkea ei tarvitse tehdä kerralla. Tietojen avaamisessa voidaan edetä vaiheittain, tekemällä oppien ja siirtymällä yksinkertaisista aineistoista kohti monimutkaisempia. Kartoituksen myötä voidaan alkaa rakentaa esimerkiksi Museoviraston datapolitiikkaa tai tiedon avaamisen strategiaa, ja selventää tavoitteita datan avaamisen hyödyntämiseksi. Tästä voi olla seurauksena esimerkiksi tietojen uudenlaiset käyttötarkoitukset, tietojen yhdistelemisen pohjalta syntyvä uusi tieto, uudenlaiset yhteistyökumppanit ja jopa organisaation roolin jalostuminen.¹⁵

Museovirasto on alansa asiantuntijaviranomainen, joten organisaation sisältä on löydettävissä myös sen hallinnoiman tiedon paras asiantuntijuus. Tietovarantojen avaamisessa kannattaa luottaa niin organisaation sisäiseen yhteistyöhön kuin mahdollisiin ulkopuolisiin yhteistyökumppaneihinkin.

*Sisäisesti työ on hyvä toteuttaa **sisällöllisten** (tietojen hallinnoinnista, keräämisestä tai tuottamisesta vastaavan osaston tai yksikön edustaja tai edustajat), **teknisten** (arkisto- ja tietopalveluiden edustaja) ja **lainopillisten** (laki- ja henkilöstöpalveluiden lakimies) **asiantuntijoiden yhteistyönä**. Näin varmistetaan, että avattavaksi suunniteltu **tieto on asianmukaista, asianmukaisessa muodossa ja lainsäädännön puitteissa julkaistavissa**.*

Museoviraston tietoaineistoja avattaessa olisi hyvä miettiä ainakin seuraavia asioita:

- Minkälaista tietoa hallinnoimme? Voidaanko joidenkin tietoaineistojen avaamista pitää kiireellisempänä kuin muiden ja millä perusteella?
- Mihin hallinnoimaamme tietoa olisi mahdollista käyttää?
- Keitä tietojen avaaminen hyödyttää? Kuka olisi kiinnostunut yhteistyöstä tietojen avaamiseen liittyen?
- Missä muodossa tieto olisi tarkoituksenmukaista avata? Onko rajapintojen rakentaminen tarpeen?
- Mikä on aineiston elinkaari? Kun tiedot on avattu, ketkä vastaavat sen ylläpitämisestä ja päivittämisestä? Kuinka usein tietoja päivitetään?

Viraston ulkopuolisten tahojen kanssa tehdystä yhteistyöstä voi olla hyötyä erityisesti tietojen avaamista suunniteltaessa ja tietojen hyödyntämiseen rohkaisemisessa. Niin sisäisessä työryhmässä kuin yhteistyökumppaneidenkin kanssa olisi hyvä miettiä esimerkiksi seuraavia asioita:

¹⁵ Poikola & Kola & Hintikka 2010, 6–7.

- Minkälaista tietoa on hyödyllistä avata ja mihin tarpeeseen avattava tieto vastaa?
- Missä muodossa tiedot ovat parhaiten hyödynnettävissä käyttäjien näkökulmasta?
- Kuinka tietojen käyttöä voisi edesauttaa?
- Minkälaista tietoa datan hyödyntäminen vaatii? (vrt. esimerkiksi [Museotilaston Tietotuoteseloste](#))

Museoviraston hallinnoimat tietoaineistot

Museoviraston hallinnoimat tietoaineistot voidaan jaotella esimerkiksi seuraavasti:

- Kirjastoaineistot
- Arkistoaineistot
- Suojelukohdetiedot
- Kokoelmat
- Mittaus- ja kartoitusaineistot
- Tutkimus- ja analyysiaineistot
- Tilastoaineistot

Datan avaajan tarkistuslista

Kun avattava tietoaineisto on valittu, täytyy määritellä avoimena datana julkaistavan tiedon laajuus (kokonaisuus, joka avataan). Tähän voivat vaikuttaa esimerkiksi aineiston käytön mahdolliset oikeudelliset rajoitteet. Nämä kannattaa tarvittaessa selvittää yhdessä lakimiehen kanssa:

- Onko aineisto julkista? Voidaanko aineisto julkaista joidenkin muutosten jälkeen?
- Sisältääkö aineisto henkilötietoja?
- Onko aineistossa tekijänoikeudella suojattua aineistoa?
- Onko avaamiseen muita rajoitteita, mm. sopimusten johdosta?

Tietojen avaamisen tukena voi käyttää esimerkiksi "[Kohti avointa kulttuuria](#)" oppaan sivulla 32 esiteltyä Kulttuuriaineistojen avaajan muistilistaa. Muistilistaa voi soveltaa niin julkishallinnollisen kuin kulttuuridatan avaamisessa:

- ✓ Avattava aineisto valittu.
- ✓ Onko aineisto (sisältö ja/tai data) avattavissa ja julkaistavissa?
- ✓ Voidaanko aineisto julkaista joidenkin muutosten jälkeen?
- ✓ Avauksen tavoitteet kirjattu.
- ✓ Avoin lisenssi valittu.
- ✓ Avoin ja koneluettava dataformaatti valittu.
- ✓ Aineiston laatu tarkistettu ja korjattu (esim. kuvan resoluutio), data siivottu ja virheet korjattu.
- ✓ Yksityiskohtainen aineiston ja datan kuvaus yhteystiedoilla tehty hyödyntäjiä varten (vrt. esimerkiksi Museotilaston Tietotuoteseloste).
- ✓ Aineiston ensisijainen sijoituspaikka valittu.
- ✓ Mahdolliset muut aineiston sijoituspaikat valittu.
- ✓ Aineiston päivityksen aikataulu ja toteutus suunniteltu.
- ✓ Avatun aineiston viestintäsuunnitelma tehty.
- ✓ Aineisto julkaistu.

Datan avaamiseen liittyvä lainsäädäntö

Julkishallinnon avoimen datan lisenssin (käyttölupa) tarkoituksena on, että käyttölupa kohteena olevaa dataa voi käyttää vapaasti kyseisen luvan ehtojen mukaisesti. Käyttölupa antaa oikeudellisen pohjan aineiston käytölle.¹⁶ Tämän luvan myöntäminen edellyttää, että datan avaajalla on oltava oikeus avata aineisto kyseiseen käyttöön.

Aineiston avaamista koskevat oikeudelliset kysymykset liittyvät tyypillisesti ainakin aineiston julkisuuteen, henkilötietojen suojaan, tekijänoikeuksiin sekä mahdollisesti eräisiin muihin erityiskysymyksiin aineiston sisällöstä riippuen, joita käsitellään alla tarkemmin.

Käyttölupa myöntäminen aineistoon ei vaikuta aineistoon itseensä, esimerkiksi luettelon, tietokannan tai muun yksittäisen teoksen tekijänoikeuteen tai muuhun oikeustilaan sellaisenaan: esimerkiksi tekijänoikeudenhaltijalla säilyy kuitenkin aina tekijänoikeus aineistonsa, mutta käyttölupa myöntämällä hän myöntää aineistonsa käyttöoikeuden kolmansille tahoille. Lisenssin ehtojen perusteella määräytyy esimerkiksi se, miten laajasti, mihin tarkoitukseen ja millä tavalla aineistoa voi käyttää. Käyttölupa määrittelee sallitun käytön rajat.

Oikeudellisesti huomionarvoista on, että kerran avattua dataa ei voida jälkikäteen enää poistaa. Näin ollen oikeus tietojen julkaisemiseen on selvitettävä huolellisesti jo ennen datan avaamista ja määriteltävä etukäteen aineisto, joka julkaistaan. Aineistoa avoimena datana avattaessa aineisto kannattaa käydä läpi, vaikka se olisikin julkaistu jo aiemmin esim. Internetissä tai painetuissa julkaisuissa, koska lainsäädäntö muuttuu jatkuvasti. Usein aineisto on mahdollista avata ainakin osittain siten, että juridisilta epäselvyyksiltä vältytään. Lainsäädännön vaikutusten selvittäminen kannattaa tehdä jo hyvissä ajoin aineistoa valittaessa.

Alla on käsitelty lyhyesti tärkeimpiä lainkohtia, jotka tulee selvittää datan avaamisen yhteydessä.

Julkisuus

Julkisuus on viranomaistoiminnassa vahva pääsääntö, jota viranomaisten tulee pyrkiä edistämään. Myös hyvä hallinto edellyttää viranomaistoiminnan läpinäkyvyyttä.¹⁷ Avattavan aineiston täytyy kuitenkin olla julkista, jotta se voidaan julkaista avoimena datana.¹⁸ Aineistoon voi liittyä sellaisia julkisuuslaissa tai

¹⁶ JHS 189 Avoimen tietoaineiston käyttölupa

¹⁷ Ks. esimerkiksi Laki viranomaistoiminnan julkisuudesta (21.5.1999/621) 1 §, 3 § ja 20 §

¹⁸ Helsingin kaupunki, hallintokeskus, Kaisa Mäkinen: Avoin data ja lainsäädäntö – mitä voin julkaista? 28.9.2012

muualla lainsäädännössä salassa pidettäväksi määriteltyjä tietoja, joita ei voida julkaista vapaasti. Julkisuuslain 24 §:ssä on määritelty tiedot, jotka ovat salassa pidettäviä. Tällaisia tietoja ovat muun muassa liike- ja ammattisalaisuudet sekä eräät maanpuolustukseen tai yksityiseen henkilöön liittyvät tiedot. Myös tilastoaineistojen salassapidosta on säädetty erikseen. Tilastoaineistojen avaamista harkittaessa julkaistavan aineiston laajuus ja julkisuus on aina selvitettävä erikseen.

Henkilötiedot

Julkinenkin asiakirja voi sisältää henkilötietoja, jonka takia henkilötietojen suoja on tutkittava erikseen. Henkilötiedon käsite on hyvin laaja. Henkilötiedolla tarkoitetaan kaikenlaisia luonnollista henkilöä taikka hänen ominaisuuksiaan tai elinolosuhteitaan kuvaavia merkintöjä, jotka voidaan tunnistaa häntä tai hänen perhettään tai hänen kanssaan yhteisessä taloudessa eläviä koskeviksi.¹⁹ Käsite kattaa siis itsestään selvästi yksityiset asiat (esim. henkilön taloudelliseen tai terveydelliseen tilaan liittyvät tiedot), jotka ovat yleensä myös arkaluonteisia, mutta toisaalta myös suhteellisen yleisluontoiset asiat, kuten ammatin tai sähköpostiosoitteen.²⁰ Subjektiiivinen arviointi sen suhteen, minkä tiedon henkilö kokee yksityisyyttään loukkaavaksi, ei ole riittävää, vaan lainsäädäntöä tulee noudattaa aina kun aineisto sisältää henkilötietoja.

Mikäli tietoa ei voi yhdistää henkilöön, kyseessä ei ole henkilötieto. Sen määrittämiseksi onko henkilö tunnistettavissa, on otettava huomioon kaikki kohtuullisesti toteutettavissa olevat keinot, joita rekisterinpitäjä voi kyseisen henkilön tunnistamiseksi käyttää. Henkilö voi olla tunnistettavissa tiedon perusteella joko suoraan tai epäsuorasti. Suorasta tunnistamisesta on kyse esimerkiksi silloin kun henkilötietona on nimi. Epäsuoralla tunnistamisella viitataan yleensä pienten tai suurten ainutkertaisten yhdistelmien ilmiöön, jotka yhdistettynä voivat johtaa henkilön tunnistamiseen. Henkilö siis saattaa silti olla tunnistettavissa, jos tieto yhdistettynä muihin tietoihin (joiden ei välttämättä tarvitse olla rekisterinpitäjän hallussa) mahdollistaa henkilön tunnistamisen.²¹

Laajaa aineistoa avattaessa riski yksityisen luonnollisen henkilön epäsuorasta tunnistamisesta unohtuu helposti, minkä takia tähän kannattaa kiinnittää erityisesti huomiota.

¹⁹ Henkilötietolaki 3.1.1 §

²⁰ Tietosuojatyöryhmä, Lausunto 4/2007 henkilötietojen käsitteestä, 20.6.2007

²¹ Tietosuojatyöryhmä, Lausunto 4/2007 henkilötietojen käsitteestä, 20.6.2007

Henkilötietoja sisältävää aineistoa ei yleensä ole mahdollista julkaista avoimena datana, ellei julkaisuun ole lain mukaista perusteltua syytä. Jos tiedot kuitenkin voidaan anonymisoida siten, että henkilö ei ole tunnistettavissa, aineisto on mahdollista julkaista.²²

Tekijänoikeus

Tekijänoikeudella suojataan itsenäisen ja omaperäisen luomistyön tulosta, jonka luonnollinen henkilö on luonut tai luonnolliset henkilöt ovat yhdessä luoneet.²³ Tätä kutsutaan teokseksi. Teos voi olla kirjallinen teos, valokuva, taideteos tai mikä tahansa muukin tuote, joka ylittää teostasoon. Laki antaa suojaa sekä koko teokselle että sen osille. Dataa avattaessa tulisikin kiinnittää huomiota siihen, sisältääkö materiaali tekijänoikeudella suojattua aineistoa. Tekijänoikeus voi ylittyä esimerkiksi kirjallisissa tuotteissa jo suhteellisen matalallakin tasolla: lähes kaikki kirjallisessa ilmenemismuodossa olevat luomistyön tuotteet saavat tekijänoikeussuojaa, jos siitä ilmenee laatijan itsenäisyys ja omaperäisyys. Informaatio tai tieto ei sellaisenaan ole tekijänoikeussuojan kohteena, vaan suoja kohdistuu sen ilmenemismuotoon eli esimerkiksi yksittäisiin sanavalintoihin.²⁴ Jo yhden tai kahden lauseen pituiset virkkeet voivat saada tekijänoikeussuojaa. Dataa avattaessa erityisesti vapaiden tekstikenttien ja valokuvien osalta on muistettava selvittää tekijänoikeudelliset rajoitukset aineiston julkaisuun. Tekijänoikeuskynnys voi ylittyä myös esimerkiksi aineiston metatiedoissakin.

Muut mahdolliset rajoitteet

Datan avaamisessa on huomattava myös mahdolliset sopimusrajoitteet, jotka aineistoon liittyvät. Jos aineisto on saatu sopimukseen perustuen, merkitystä on myös sillä, millaiset oikeudet aineiston käyttöön on saatu. Joskus aineiston julkaisemisesta tai uudelleen käytöstä on tehty määräyksiä, joita tulee noudattaa. Kyse voi olla esimerkiksi yksityisen toimijan kanssa tehdystä kirjalliseen työhön liittyvästä toimeksiantosopimuksesta, jonka yhteydessä on rajoitettu mahdollisuutta luovuttaa aineistoa eteenpäin.²⁵ Tällaista aineistoa ei voida avata avoimena datana.

Datan avaamiseen voi liittyä myös muita juridisia kysymyksiä, joita kannattaa selvittää lakimiehen kanssa.

²² Helsingin kaupunki, hallintokeskus, Kaisa Mäkinen: Avoin data ja lainsäädäntö – mitä voin julkaista? 28.9.2012

²³ Tekijänoikeuslaki (8.7.1961/404) 1 §

²⁴ Harenko-Niiranen-Tarkela: Tekijänoikeus, kommentaari ja käsikirja, 2006

²⁵ Helsingin kaupunki, hallintokeskus, Kaisa Mäkinen: Avoin data ja lainsäädäntö – mitä voin julkaista? 28.9.2012

Aineiston saattaminen julkaistavaan muotoon

Iso osa tiedon avaamiseen liittyvää työtä on valitun aineiston muokkaaminen julkaistavaan muotoon.

Tämä voi pitää sisällään tietoa-aineiston siivoamista, mahdollisesti salattavien aineiston osien poistamista tai piilottamista sekä aineiston muokkaamista käsiteltävään muotoon. Tähän työhön on hyvä varata aikaa ja siinä kannattaa olla mukana niin aineistoon perehtynyt asiantuntija kuin tiedonhallinnan asiantuntijakin.

Avattavat tiedot voivat olla alun perin erilaisissa muodoissa esim. erilaisia taulukoita tai tietokantoja, kuvia tai asiakirjatiedostoja. Taulukot voivat sisältää itsessään olennaista sisältöä tai niissä voi olla esim.

metatietoa joka kuvailee tietoon liittyvien tiedostojen luonnetta ja sisältöä. Jos tietosisältö on avattavaksi sopivaa ja siitä saa muodoltaan järkevän excel-tilukon on avaaminen jo hyvällä mallilla. Lähtöaineiston käsitteleminen ja koostaminen voi olla vaativaa ja on hyvä pyytää apuun excel-osaaja jo varhaisessa vaiheessa. Teknisissä kysymyksissä voi hakea tukea myös Kansallisen digitaalisen kirjaston KDK-standardisalkusta²⁶.

Taulukkomuotoisen tiedon pitäisi olla rakenteeltaan mahdollisimman yksinkertaista. Koneluettavaa tietoa tavoiteltaessa monimutkaiset otsikoinnit, yhdistellyt solut ja visuaaliset tehosteet ja muu erikoinen muotoilu tuo yleensä ongelmia. Ihmisenkin on helpompi tulkita uutta tietoa kun se on yksikertaisessa taulukossa. Suositeltavaa olisikin että taulukossa olisi selkeät otsikot, kukin tieto omassa solussaan ja tulkittavissa ilman sisällön lihavoiteja, värillisyyttä tai muita tehosteita.

Toinen tärkeä seikka on tietojen yhdenmukaisuuden varmistaminen. Jos jokin sarake sisältää numerotietoa on hyvä varmistaa, että kaikki tiedot myös ovat numeroita. Samoin vakioarvojen pitää olla aina samalla tavalla esim. kuvitteellisessa kentässä "aukioloaika" ei saisi olla "vuorokauden ympäri", "24h" ja "24 tuntia vuorokaudessa" jotka kaikki tarkoittavat samaa asiaa. Aina ei ole mahdollista olla täysin yhdenmukainen, mutta tällaisissa tapauksissa asiasta pitää kertoa tiedon käyttäjälle esimerkiksi tietotuoteselosteessa.

CSV-tiedostomuoto, jossa avointa dataa usein jaetaan asettaa myös tiettyjä ehtoja. Vaikka CSV (comma separated values) viittaakin sarakkeiden ja tietojen erottamiseen toisistaan pilkulla, käytetään erottimena yleisimmin puolipistettä. Pilkun käyttö erottimena tuo kotimaisissa aineistoissa ongelmia, koska se ei erotu desimaalipilkusta. Puolipistettä käytettäessä pitää huolehtia että taulukossa mahdollisesti olevassa tekstisisällössä ei ole käytetty puolipisteitä. On myös syytä muistaa että CSV-tiedosto on tekstitiedosto joka ei sisällä mitään muotoiluja. Esim. sarakelevydet yms. eivät siirry tiedostojen mukana.

²⁶ <http://www.kdk.fi/fi/kokonaisarkkitehtuuri/standardisalkku>

Kuvatiedostojen ja dokumenttien avaamisessa kannattaa käyttää tiedostomuotoja jotka ovat helposti mahdollisimman monien käyttäjien hyödynnettävissä. KDK:n pitkäaikaissäilytys on listannut säilytys ja siirtokelpoisia tiedostomuotoja²⁷ ja listaa kannattaa hyödyntää myös dataa avattaessa. Kuvatiedostojen osalta suosituimmat ja yhteensopivimmat lienevät jpg ja png. Dokumenttien ja asiakirjojen tiedostomuotona PDF lienee suosituin. Se säilyttää tekstisisällön, kuvat ja muotoilut ja on laajalti käytössä. Käyttäjän kannalta PDF-tiedostossa ongelmallista voi olla se, että tietosisällön jatkokäyttö ja muokkaaminen on usein hankalaa. PDF-tiedostosta on olemassa myös ns. PDF-A-versio, joka on arkistokäyttöön hyväksytty sähköinen säilytysformaatti.

Yllä mainitut tiedostomuodot ovat ihmisten luettavissa ja osittain myös koneluettavia. Pelkästään koneluettavat tiedostomuodot kuten XML tai RDF ja näitä tarjoavat koneluettavat rajapinnat ovat teknisesti huomattavasti haastavampia toteuttaa. Tällaista toteutusta suunniteltaessa voisi suositella että ensin tutustuttaisiin muiden toimijoiden tarjoamiin aineistoihin ja niiden käyttämiseen, jotta oman aineiston käyttömahdollisuudet hahmottuisivat paremmin.

Tietotuoteseloste ja sen laadinta

Avattavasta tietoaaineistosta kannattaa aina tuottaa niin kutsuttu "tietotuoteseloste". Tietotuoteselosteessa esitellään avattu aineisto ja siihen liittyvää, käyttäjälle oleellista tietoa esimerkiksi aineiston keruuseen, aineiston luotettavuuteen, käytettyihin käsitteisiin, tiedon rakenteeseen tai odotettaviin käyttötarkoituksiin liittyen. Tietotuoteselosteen sisältö kannattaa aina määritellä avatun tietoaaineiston pohjalta, mutta siihen on saatavilla esimerkkejä jo avattujen aineistojen yhteydessä.

On myös tärkeää, että tietotuoteselosteesta käy ilmi, kuinka usein ja milloin tietoja päivitetään sekä tiedon ylläpidosta vastaavan tahon yhteystiedot mahdollista palautetta ja kysymyksiä varten.

Tietotuoteseloste kannattaa lisensoida joko samalla lisenssillä kuin itse aineisto tai mahdollisuuksien mukaan vapauttaa se täysin vapaaseen käyttöön [CC0-lisenssillä](#).

Tietoaaineistojen ylläpito ja päivittäminen

Tietoaaineistojen avaamisen jälkeen tietoa ja siihen liittyvää tietotuoteselostetta on hyvä päivittää saadun palautteen pohjalta. Mikäli kyseessä on karttuva aineisto, on palaute hyvä ottaa huomioon vähintään tietoja kartutettaessa ja uusia osoita avattaessa. Palautteen vastaanotossa ja hyödyntämisessä kannattaa yhdistää sisällöllinen ja tekninen osaaminen tarvittavin osin.

²⁷ <http://www.kdk.fi/images/tiedostot/KDK-PAS-tiedostomuodot-v1.3.pdf>

Tietoaineistojen käytön edistäminen

Jo tietojen avaamista suunniteltaessa on hyvä miettiä erilaisia tapoja tiedon käytön lisäämiseksi.

Julkishallinnollinen data kannattaa vähintään viedä avoindata.fi -portaaliin, mutta tämän lisäksi voi miettiä yhteistyötä esimerkiksi avoimen datan hyödyntämiseen keskittyvien verkostojen tai muiden datan potentiaalisten käyttäjien kanssa. Ohjeita tähän löytyy esimerkiksi AvoinGLAMin tekemästä oppaasta "[Kohti avointa kulttuuria](#)".

Lisäksi kannattaa mahdollisuuksien mukaan tarjota aineistoja käytettäväksi esimerkiksi [Apps4Finlandin](#) kaltaisiin kilpailuihin, tai järjestää itse hackathon yhdessä kentän muiden toimijoiden kanssa.

Case: Museotilasto

Museoviraston tehtävänä on suomalaista museoalaa koskevan toimialatilaston kerääminen. Tilastoa kerätään verkkopalvelun avulla vuosittain ammatillisesti hoidetuilta museoilta. Samaan aikaan tietojen avaamiseen tähtäävän projektin kanssa oli käynnissä myös tilastoinnissa käytetyn verkkopalvelun uudistaminen.

Museotilastoa on kerätty sähköisesti vuoden 2007 tiedoista lähtien. Sähköisen palvelun vahvuutena voidaan pitää tiedon keräämisen tehokkuutta, mutta ennen kaikkea tiedon saavutettavuuden paranemista. Syksyn 2014 ja kevään 2015 aikana tilastoinnin verkkopalvelua uudistettiin ja erityistä huomiota kiinnitettiin nimenomaan tiedon saavutettavuuteen mm. palvelun hakutoiminnallisuuksia kehittämällä. Samalla myös vuodesta 2007 lähtien kerätyt tiedot siirretään selattaviksi uuden verkkopalvelun kautta. Tämän työn luonnolliseksi jatkoksi nähtiin tilastotiedon lisensointi eli käyttöluvan määrittely sekä tiedon julkaiseminen avoimena datana.

Lisensoinnin avulla voidaan selventää tilastotiedon käyttöoikeuksia, mikä samalla helpottaa tiedon käyttöä. [JHS-suosituksen](#) mukaisesti Museotilastossa kerättyyn tietoon päätettiin soveltaa avointa lisenssiä *Creative Commons Nimeä 4.0 Kansainvälinen (CC BY 4.0)*.

Museotilastossa kerätyn tiedon osalta valittu lisenssi tarkoittaa, että tiedon käyttöä, jakelua ja muokkausta ei rajoiteta millään tavalla, mutta käyttäjän tulee viitata alkuperäiseen tietolähteeseen, eli Museoviraston tuottamaan Museotilastoon, ja merkittävä, mikäli aineistoon on tehty muutoksia. Näin turvataan loppukäyttäjän mahdollisuudet alkuperäisen tiedon löytämiseen sekä jaetun tiedon oikeellisuuden tarkistamiseen.

Työn edetessä todettiin, että data voidaan julkaista, mikäli se ei sisällä yksityisyydensuojan piiriin kuuluvaa tietoa, kuten henkilötietoja tai aineistoa, joka tulkitaan arkaluontoiseksi tai salassa pidettäväksi, kuten kansalliseen turvallisuuteen liittyviä aineistoja. Tarkoituksena oli julkaista avoimena datana se osa tilastokyselyssä kerätystä tiedosta, joka on jo aikaisemmin ollut www.museotilasto.fi-sivuston kautta ladattavissa. Avattava data ei siis sisällä esim. museoiden yhteystietoja tai avoimia vastauksia, vaan pääasiassa numeerisia vastauksia.

Vuosittaiset tilastot ovat haettavissa Museotilasto.fi-sivuilta kuten tähänkin asti. Tietojen käyttäjille kerrotaan sivustolla käyttöehdot, mikä tekee tilastotietojen hyödyntämisestä aikaisempaa selkeämpää. Vuosittaiset tilastot laitettiin saataville veloitusetta Museotilasto.fi-palveluun tavallisina Excel-taulukoina, mutta myös koneluettavassa csv-muodossa. Valitun toteutustavan ansiosta tilastodata voidaan määritellä kolmen tähden avoimeksi dataksi. Avattu aineistopaketti tietotuoteselosteineen päätettiin listata löydettävyyden parantamiseksi ja käytön edistämiseksi myös avoindata.fi-portaaliin.

Valittu lisenssi tuodaan selvästi esiin Museotilasto.fi-palvelussa sekä aineiston yhteydessä. Lisensointi koskee jo kerättyä tietoa (vuoteen 2013 saakka), mutta lisensoinnin jälkeen se tulee koskemaan myös vuosittain julkaistavaa uutta aineistoa.

Tiedon selkeällä lisensoinnilla helpotetaan tiedon käyttöä, siihen viittaamista ja tiedon jakamista.

Tilastotiedon avaamisella odotetaan saatavan uudenlaista käyttöä tilastotiedolle, mutta myös uutta tietoa niin museoiden kuin viranomaistenkin käyttöön. Tilastotiedon avaaminen mahdollistaa tiedon yhdistämisen muihin tietoihin, jolloin voidaan saada uudenlaista tietoa esimerkiksi museoiden alueellisesta merkityksestä.

Samalla tietojen avaaminen antaa Museovirastolle ja museoille lisää kokemusta tietojen avaamisesta, siihen liittyvästä osaamisesta ja vaadittavista resursseista.

Hankkeen taustat Museovirastossa ja tiedon avaamisen tavoitteet

Tietojen avaaminen noudattaa julkisen hallinnon datan avaamisen yleisiä linjauksia (mm. PSI-direktiivi ja [Avoimen tiedon ohjelma, VM](#)). Opetus- ja kulttuuriministeriö kannustaa hallinonalansa viranomaistoimijoita tuottamansa tiedon avoimeen julkaisuun.

Museotilastossa kerätty tieto haluttiin nyt julkaista avoimena datana, sillä kerätty tieto on kompaktilta ja karkuu tasaisesti vuosittain. Museotilaston vastausprosentti on lähes 100, mikä tekee siitä erittäin arvokkaan tietovarannon Suomen ammatillisesti hoidetusta museokentästä. Tilastotiedon avaamisella odotetaan saatavan uudenlaista käyttöä tilastotiedolle, mutta myös uutta tietoa niin museoiden kuin

viranomaistenkin käyttöön. Tilastotiedon avaaminen mahdollistaa tiedon yhdistämisen muihin tietoihin, jolloin voidaan saada tietoa esimerkiksi museoiden alueellisesta merkityksestä.

Ennen kevättä 2015 Museotilastossa kerättyjen tietojen käyttöoikeuksia ei ollut erikseen määritelty, mikä hankaloitti osaltaan tiedon käyttöä ja siihen viittaamista. **Tiedon selkeällä lisensoinnilla voidaan edesauttaa ja helpottaa tiedon käyttöä, siihen viittaamista ja tiedon jakamista.** Samalla projekti antoi Museovirastolle lisää kokemusta tietojen avaamisesta, siihen liittyvästä osaamisesta ja vaadittavista resursseista.

Julkisen hallinnon tietohallinnon neuvottelukunnan (JUHTA) sekä sen alaisen JHS-jaoston 11.12.2014 hyväksytyn suosituksen mukaisesti Museotilastossa kerättyyn tietoon myönnettiin käyttöoikeudet avoimella käyttöluvalla **Creative Commons Nimeä 4.0** (CC BY 4.0

<https://creativecommons.org/licenses/by/4.0/deed.fi>). JHS-suositus: <http://www.jhs-suositukset.fi/web/guest/jhs/recommendations/189>.

Tilastodatan avaamiseen liittyvä lainsäädäntö

Kuten edellä on tuotu esille, lainsäädäntö asettaa rajoituksia sille, kuinka laajasti aineistoa voidaan julkaista avoimena datana. Museotilaston tapauksessa oikeudelliset kysymykset koskivat erityisesti tilastolain tulkintaa sekä henkilötietolakia. Aineisto sisälsi pääasiassa vain kyllä/ei-vastauksia ja numeraalista dataa, jonka takia esimerkiksi tekijänoikeudelliset kysymykset eivät tulleet merkittäväksi ongelmaksi. Selvitykset dataan liittyen koskivat sen sijaan erityisesti julkisuuslain ja tilastolain välistä suhdetta sekä yksittäisen luonnollisen henkilön tunnistettavuuden mahdollisuuden estämistä.

Museotilaston tapauksessa selvittäminen aloitettiin jo hyvissä ajoin datan avaamista suunnitellessa, mikä mahdollisti aineiston kattavan läpikäynnin. Aineisto oli jo aiemmin avattu Internetissä, mutta se haluttiin kertaalleen käydä läpi vielä ennen datan avaamista avoimena datana. Selvityksen yhteydessä data käytiin läpi ja siitä poistettiin sellainen aineisto, jonka julkaisemista ei esimerkiksi lainsäädännössä asetettujen rajoitusten johdosta katsottu tarpeelliseksi. Aineistoa rajoitettiin ja anonymisoitiin myös sen varmistamiseksi, että yksityisydensuoja voidaan turvata kaikissa tilanteissa. Tämän johdosta aineistosta esimerkiksi piilotettiin eräät vastaukset, joiden otos oli alle 5. Lisäksi joitain kysymyksiä on poistettu kokonaan julkaistavien joukosta.

Avaamisen vaiheet

Tietojen avaaminen voi aineiston laajuudesta riippuen olla pitkäkin prosessi, mutta sen voi jakaa vaiheisiin. Museotilaston tietoja avattaessa ensimmäinen vaihe oli aineiston valinta ja työryhmän muodostaminen. Tämän jälkeen avattavaksi suunniteltu aineisto käytiin läpi niin sisällön kuin lainsäädännönkin näkökulmasta. Ennalta on vaikea arvioida, minkälainen tietosisältö kiinnostaa eniten, minkä vuoksi kannattaa ensisijaisesti kiinnittää huomiota siihen, että julkaistava tieto on mahdollisimman oikeellista ja lainsäädännön kannalta julkaistavissa. Myöhemmin saadun palautteen pohjalta tietoaineiston rakennetta ja sisältöä voi tarpeen tullen muokata. Samaan aikaan tietojen läpikäynnin kanssa toteutettiin tilastoinnin piiriin kuuluville museoille suunnattu kommentointikierros, jonka aikana pyrittiin antamaan mahdollisimman tarkkaa tietoa tiedon avaamisesta, valitusta lisenssistä ja avattavasta tietoaineistosta. Valitun aineiston ja saatujen kommenttien perusteella valmisteltiin vielä aineistoa kuvaava tietotuoteseloste. Lopuksi avattavat aineistot ja tietotuoteseloste vietiin museotilasto.fi-verkkopalveluun ja avoindata.fi -portaaliin, missä ne ovat vapaasti kaikkien saatavissa.

Aineiston valinta ja sisäisen työryhmän muodostaminen

Tilastoaineisto vuodesta 2007 lähtien oli alun alkaen sähköisessä muodossa, minkä vuoksi sen avaamiselle koettiin olevan hyvät edellytykset. Hyvin aikaisessa vaiheessa aineiston käsittelyä tuli selväksi, että projektin kannalta olisi tärkeää tehdä yhteistyötä Museoviraston eri asiantuntijoiden välillä, jotta niin lainopilliset kuin teknisetkin näkökulmat voitaisiin ottaa alusta asti huomioon tietojen käsittelyssä. Työryhmään kuului kaksi aineiston sisältöön perehtynyttä asiantuntijaa, lakimies sekä tiedonhallinnan asiantuntija. Työn edetessä aineistoa karsittiin ja muokattiin julkaistavaan muotoon.

Koska aineisto päätettiin julkaista excel- ja csv-muodossa, oli tekninen tuki myös tarpeen. Excel-taulukoissa jo valmiina olevaa tietosisältöä tarkistettiin ja välimerkkien käyttöä yhdenmukaistettiin koneluettavuuden helpottamiseksi. Tietotuoteselosteessa on tuotu tähän liittyen ilmi kohtia, jotka vaikuttavat tietojen käyttöön (ks. [Museotilaston tietotuoteseloste](#)).

Julkaistavien tietojen laajuus

Projektin aikana tehtyjen selvitysten pohjalta voitiin todeta, että museotilastossa kerätyt tiedot voidaan julkaista, mikäli ne eivät sisällä esimerkiksi yksityisyydensuojan piiriin kuuluvaa tietoa, kuten henkilötietoja, tai aineistoa, joka tulkitaan arkaluontoiseksi tai salassa pidettäväksi, kuten kansalliseen turvallisuuteen liittyviä aineistoja.

Tarkoituksena oli julkaista avoimena datana se osa tilastokyselyssä kerätystä tiedosta, joka on jo aikaisemmin ollut www.museotilasto.fi-sivuston kautta ladattavissa ja käytettävissä. Avattava data ei siis

sisältäisi esim. museoiden tarkkoja yhteystietoja tai avoimia vastauksia, vaan pääasiassa numeerisia vastauksia.

Museoille suunnattu kommentointikierros

Vaikka Museovirastolla on oikeudet Museotilaston tietojen avaamiseen haluttiin tilastoon tietoja antavat museot ottaa mukaan prosessiin ja antaa kaikille mahdollisuus kommentoida, keskustella ja saada halutessaan lisätietoa asiaan liittyen.

Museoille suunnattu kommentointikierros käynnistettiin 17.12.2014. Lähetetyssä viestissä esiteltiin ehdotettu käyttöluja ja kuvattiin valinnan taustat, esiteltiin avattavaksi aiottu tietoaineisto sekä kuvattiin tietojen avaamisen tavoitteita ja käytännön toteutustapaa. Museotilastoon vastaavia museoita pyydettiin kommentoimaan tilastotietojen lisensointia ja julkaisua avoimena datana 16.1.2015 mennessä.

Kommentteja saatiin määräaikaan mennessä kuudelta museolta – kaikissa kommenteissa tilastotiedon avaaminen nähtiin positiivisena asiana, joka voi parhaimmillaan lisätä museoalaan liittyvää tietoa sekä museoiden tunnettuutta. Lisäksi ehdotettu käyttöluja (CC BY 4.0) ja julkaistavien tietojen rajausta koettiin pääsääntöisesti sopiviksi. Lisätietoja kaivattiin seuraavista aiheista:

- Hankkeen taustat Museovirastossa ja tiedon avaamisen tavoitteet
- Tilastotietojen käytettävyyteen liittyvät haasteet ja numeraalisen datan taustoihin perehtyminen käytön edellytyksenä
- Julkaistavien tietojen laajuus ja mahdollisuus suppeamman kokonaisuuden julkaisuun
- Suostumus tietojen julkaisuun
- Käytännön toteutus

Museoille lähetettiin kommenttien pohjalta uusi viesti, jossa pyrittiin avaamaan lisää kohtia, jotka olivat herättäneet kysymyksiä.

Tietotuoteseloste ja sen sisältö

Kommentointikierroksen aikana nostettiin esiin, että tilastotiedon käyttö ja vertailu vaatii perehtymistä numeraalisen datan taustoihin ja keruussa käytettyihin käsitteisiin.

Museotilastotiedon kuten minkä tahansa tilastotiedon käytön edellytyksenä on aina perehtyneisyys kyseessä olevan tiedon taustoihin. Museotilastotietojen lisensoinnin ja tilastohakupalvelun kehittämisen yhtenä osa-alueena oli mm. Museotilastossa käytettyjen käsitteiden määrittely ja niitä koskevan tiedon löydettävyyden parantaminen.

Kun tilastotieto julkaistiin avoimena datana, julkaistiin myös niin sanottu tietotuoteseloste (ks. [Museotilaston tietotuoteseloste](#)), joka voidaan nähdä tilastotietoja koskevana metatietona. Vuosittaiset

taulukot, tietotuoteseloste sekä käytettyjen käsitteiden määrittelyt tarjotaan verkkopalvelussa yhtenä pakettina.

Tietotuoteselosteen laadintaa varten on olemassa Museovirastossa valmiita pohjia, mutta myös kommentointikierröksellä esitettyjen kysymysten pohjalta selosteessa voitiin tarttua asioihin, jotka mahdollisia käyttäjiä voivat askarruttaa.

Tilastodataa ja tietotuoteselostetta on tarkoitus tarkentaa ja päivittää julkaisun jälkeen saadun palautteen pohjalta. Käytäntö on yleinen tietoa avattaessa ja kuuluu oleellisena osana avoimen tiedon ylläpitoon.

Suostumus tietojen julkaisuun

Tilastokyselyyn päätettiin vuoden 2014 tietojenkeruusta lähtien lisätä oma kohtansa tietojen käytön sallimiseksi. Suostumus koskee tietojen käyttöä toimitetussa tilastojulkaisussa, tilastotietojen julkaisua avoimena datana ja tietojen saatavuutta museotilasto.fi-palvelun Tilastohaku-osiossa.

Tiedot vuosilta 2007–2013 voitiin 16.1.2015 päättyneen kommentointikierröksen aikana saadun positiivisen palautteen ansiosta julkaista avoimena datana helmikuussa 2015.

Nähdäksemme on koko museokentän yhteinen etu, että tiedot voidaan vastaisuudessakin julkaista yhtä laajana kokonaisuutena mahdollisimman avoimesti. Tämä hyödyttää niin museotoimijoita, viranomaisia kuin muitakin tiedon käyttäjiä. Toimialatilaston pohjalta voidaan muodostaa kattavampi kokonaiskuva alan nykyhetkestä, mutta myös pitkän aikavälin trendeistä.

Käytännön toteutus

Kommentointikierröksen pohjalta

Museotilaston tiedot lisensoitiin

avoimella lisenssillä **Creative Commons Nimeä 4.0 (CC BY 4.0)**.

Tämä tarkoittaa, että tiedon

käyttöoikeudet tuodaan selvästi esiin

museotilasto.fi-palvelussa sekä julkaistun aineiston yhteydessä.

Vuosittaiset tilastot ovat saatavilla

veloituksetta museotilasto.fi-palvelussa tavallisina Excel-taulukoina, mutta myös koneluettavassa csv-muodossa yhdessä tietotuotemäärittelyn kanssa. Näin ollen tilastodata voidaan datan viiden tähden mallin mukaisesti luokitella kolmen tähden dataksi. Vuosittaiset koneluettavat tiedostot metatietoineen on

	A	B	C	D
1	Museoiden toiminta 2013			
2	Lisenssi:	Creative Commons Nimeä 4.0 Kansainvälinen (CC BY 4.0)		
3	Lisenssin url:	https://creativecommons.org/licenses/by/4.0/deed.fi		
4	Lähde:	Museotilasto 2013, Museovirasto		
5	Päivitetty:	23.2.2015		
6				
7				
8	Kysymyksen koodi	museo	museonumero	kunta

Valittu lisenssi on hyvä tuoda esiin myös aineiston yhteydessä. Museotilaston vuosittaisissa taulukoissa vasempaan yläkulmaan on lisätty lisenssiä koskevat tiedot, tietojen lähde ja tieto siitä, milloin taulukot on viimeksi päivitetty.

listattu löydettävyyden parantamiseksi ja käytön edistämiseksi myös avoindata.fi-portaaliin, jonne kerätään julkisen hallinnon tuottamaa avointa dataa.

Lopuksi

Yhteiskuntamme rakentuu yhä voimakkaammin digitaalisten tuotteiden ja palveluiden ympärille. Avoimemman yhteiskunnan rakentamisessa olennaista on tietojen saavutettavuus ja toiminnan läpinäkyvyys. Vaikka taustalla on äärimmäisen monimutkainen todellisuus, voidaan tiedon avaamista ja toiminnan muutosta viedä eteenpäin pienissä erissä, kokeilemalla ja kokemuksista oppimalla. Kuten teoksessa ”Julkinen data - johdatus tietovarantojen avaamiseen” todetaan:

Kansallisten tietovarantojen maksuttoman käytön salliminen ja sitä tukeva toimintakulttuurin muutos ovat merkittävimpiä asioita, joita suomalaisen tietoyhteiskunnan kehittämiseksi voitaisiin tällä hetkellä toteuttaa. Tämä toisi uusia mahdollisuuksia vastata ajankohtaisiin haasteisiin. Maksuton julkinen data olisi myös suomalaisille instituutioille mahdollisuus uudistua ja oppia verkostoaikakauden yhteistoimintataitoja sekä suunnata suomalaisen yhteiskunnan kehittämistä uudelleen yhdessä tekemisen kulttuuriin.²⁸

²⁸ Poikola & Kola & Hintikka 2010, 11.

Kirjallisuutta ja tietolähteitä

Käytetty kirjallisuus

Antti Poikola, Petri Kola, Kari A. Hintikka. *Julkinen data – johdatus tietovarantojen avaamiseen*. Liikenne- ja viestintäministeriö, 2010. CC BY 1.0 <http://creativecommons.org/licenses/by/1.0/fi/>.

Sanna Marttila ja Laura Sillanpää. *Kohti avointa kulttuuria*. AvoinGLAM, 2014. CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0/deed.fi>.

Tietolähteitä

"5 * Open data". <http://5stardata.info/>

"Kohti avointa kulttuuria" <http://avoinglam.fi/opas.pdf>

"Mitä on avoin data?", Helsinki Region Infoshare. <http://www.hri.fi/fi/mita-on-avoin-data/>

"Visio ja arvot", Open Knowledge Finland. <http://fi.okfn.org/about/visiojaarvot/>

Avoin hallinto <http://vm.fi/hallinnon-avoimuus/avoin-hallinto>

Creative Commons <http://creativecommons.org/>

Datan avaajan kurssi <https://p2pu.org/en/courses/2486/datan-avaaminen-johdanto/>

Digital Agenda for Europe <http://ec.europa.eu/digital-agenda/>

[Helsingin kaupunki, hallintokeskus, Kaisa Mäkinen: Avoin data ja lainsäädäntö – mitä voin julkaista? 28.9.2012](#)

Henkilötietolaki 22.4.1999/523

JHS 189 Avoimen tietoaaineiston käyttöluja <http://www.jhs-suositukset.fi/web/guest/jhs/recommendations/189>

KDK:n säilytys ja siirtokelpoiset tiedostomuodot. <http://www.kdk.fi/images/tiedostot/KDK-PAS-tiedostomuodot-v1.3.pdf>

KDK-standardisalkku <http://www.kdk.fi/fi/kokonaisarkkitehtuuri/standardisalkku>

Kristiina Harenko - Valtteri Niiranen – Pekka Tarkela: Tekijänoikeus, kommentaari ja käsikirja, 2006

Laki viranomaisten toiminnan julkisuudesta 21.5.1999/621

Open Government Partnership <http://www.opengovpartnership.org/>

Riktlinjer för digitalisering av kulturarvsmaterial på Riksantikvarieämbetet

http://www.raa.se/app/uploads/2013/10/digitaliseringsriktlinjer_riksantikvarieambetet.pdf

Tekijänoikeuslaki 8.7.1961/404

[Tietosuojatyöryhmä, Lausunto 4/2007 henkilötietojen käsitteestä, 20.6.2007](#)

Tilastolaki 23.4.2004/280

Wikimedia Suomi <http://fi.wikimedia.org/wiki/Etusivu>

Wikipedia: Wikiprojekti Kiasma/Wikimaraton

http://fi.wikipedia.org/wiki/Wikipedia:Wikiprojekti_Kiasma/Wikimaraton