



# Datan avaamisen ABC satamanpitäjille

*”Satamissa toimivat organisaatiot tuottavat ja keräävät tietoa oman tehtävänsä hoitamiseksi. Tietoja voidaan julkaista avoimena datana. Avointa dataa pystytään rikastamaan ja yhdistelemään muista tietolähteistä saatavan datan kanssa. Avoimen datan pohjalta on mahdollista kehittää uutta informaatiota, palveluja ja tehokkuutta sataman ja sen yhteydessä toimivien yritysten liiketoimintaan.”*

Jari Salo

Heikki Laaksamo

Tommy Ulmanen

Reima Helminen

Janne Saarikoski

DigiPort-hanke



Euroopan unioni  
Euroopan aluekehitysrahasto

Vipuvoimaa  
EU:lta  
2014–2020

KYMEN  
LAAKSON  
LIITTO



Uudenmaan liitto  
Nylands förbund



VARSINAIS-SUOMEN LIITTO  
Egentliga Finlands förbund  
REGIONAL COUNCIL OF SOUTHWEST FINLAND

PORT OF  
TURKU  
FINLAND



PORT OF  
HAMINAKOTKA



M E R I  
K O T K A



Kaakkois-Suomen  
ammattikorkeakoulu



MERENKULUKALAN KOULUTUS-  
JA TUTKIMUSKESKUS



TIEKE

## Sisällys

1.	Taustaa – DigiPort-hanke ja sataman infrastruktuuridatan avaaminen .....	2
2.	Datan avaamisen periaatteita .....	2
3.	Datan avaaminen- ja harmonisointi .....	2
3.1.	Datan avaamisprosessi .....	2
3.2.	Datan yhdenmukaistaminen ja standardisointi.....	5
4.	Sataman avoin data ja tiedonhallinta.....	5
5.	Standardit, datan esitystavat ja rakenteet .....	6
5.1.	Lyhyt katsaus eräisiin avoimen datan esitystapoihin .....	7
5.2.	Avoimen datan rajapinnoista ja saatavuudesta .....	8
6.	Lisensiointimallit avoimen datan tietokannoissa .....	9
6.1.	Creative Commons -lisenssit .....	9
6.1.1.	Creative Commons nimeä 4.0 .....	9
6.2.	Julkaistava data ja siihen liittyvät rajoitukset.....	10
6.3.	Tietokantojen tekijänoikeudet .....	10
6.3.1.	Kirjallisenä teoksena suojatut tietokannat.....	10
6.3.2.	Sui generis- ja luettelosuojalla suojatut tietokannat.....	10
6.3.3.	Raakadataa sisältävät tietokannat .....	11
6.4.	Lisensiointimalleista .....	11
6.4.1.	Suosittelavia lisensiointimalleja .....	12
7.	Avoimen datan varantojen ylläpidossa huomioitavia asioita.....	12
7.1.	Datan laadun ylläpito.....	12
7.2.	Datan yhtenäisyyden varmistaminen.....	13
8.	Satamien kokemuksia datan avaamisesta.....	14
8.1.	Pilottisatamien caset .....	14
8.2.	Muita esimerkkejä datan avaamisesta .....	14
8.3.	Hack the Port -kilpailu datan avaamisesta ja avoimen datan hyödyntämiskäytännöistä .....	15
8.3.1.	Hack the Port -kilpailun ratkaisuehdotuksia.....	15

# DATAN AVAAMISEN ABC SATAMANPITÄJILLE

## 1. Taustaa – DigiPort-hanke ja sataman infrastruktuuridatan avaaminen

Satamanpitäjät ja satamissa toimivat organisaatiot tuottavat ja keräävät tietoa ensisijaisesti oman tehtävänsä hoitamiseksi sekä toiminnallisten vaatimustensa ja viranomaisvaatimusten täyttämiseksi.

Kerätyt tiedot tallennetaan tietokantoihin tai tiedostoihin organisaation tietojärjestelmissä. Näitä tietokantoja ja tiedostoja on mahdollista avata avoimeksi dataksi. Avointa dataa voidaan edelleen rikastaa ja yhdistellä muista tietolähteistä saatavan datan kanssa ja kehittää tämän pohjalta uutta informaatiota ja palveluita.

Satamissa toimivat organisaatiot voivat lisätä läpinäkyvyyttään ja yhteistoimintaa kumppaneittensa kanssa, sekä mahdollisesti jakaa myös kustannuksia ja samaan aikaan parantaa datansa laatua.

DigiPort-hankkeessa<sup>1</sup> tutkitaan satamien digitalisaatiota ja kehitetään avoimen datan hyödyntämistä satamissa. Hankkeessa kannustetaan ja tuetaan sataminpitäjiä kartoittamaan omat tietovarantonsa ja luomaan uusi toimintamalli infrastruktuuriin liittyvän tiedon julkaisemiseksi avoimena datana.

## 2. Datav avaamisen periaatteita

Avoin data on digitaalista tietoa, joka on julkista, muodoltaan koneluettavaa, maksutonta ja omistajansa vapaaseen käyttöön lisensoimaa. Se on siis kenen tahansa vapaasti, laillisesti, teknisesti ja veloituksetta uudelleenkäytettävissä.

Avoimen datan hyödyntämiseksi käyttäjä tarvitsee lailliset ja selvästi ilmaistut oikeudet. Mitä laajemmat oikeudet ovat, sitä enemmän ne kannustavat tiedon hyödyntämiseen.

Avoin data ei saa sisältää tietoa, joka voidaan kohdistaa yksittäiseen henkilöön niin että hänen yksityisyydensuojansa joutuisi vaaraan. Se ei saa myöskään sisältää liikesalaisuuksiksi luokiteltua tietoa.

Avoimen datan sisältö, rakenne ja sen merkitys kuvataan metadatan avulla. Metadatalta kuvaillaan selkeästi ja ymmärrettävästi datan sisältö. Metadatalta pyritään tehostamaan tietovarannon käyttöä, koska se voi helpottaa tietojärjestelmien välisiä tiedonsiirtoja ja eri paikoissa olevien sisältöjen yhdistämistä. Täten laadukkaat metatiedot voivat myös parantaa informaation löydettävyyttä niin, että hakukoneet pystyvät etsimään tietoa tarkemmin ja monipuolisemmin. Metadata on erittäin tärkeää tietoa datan jatkojalostajalle ja auttaa käyttäjää avoimen datan oikeassa hyödyntämisessä

Avoin data ja sen hyödyntämisen mahdollistavan rajapinnan avaaminen ovat satamien innovaatiotoiminnan, kasvun ja läpinäkyvyyden kannalta keskeisiä tekijöitä.

## 3. Datav avaaminen- ja harmonisointi

### 3.1. Datav avaamisprosessi

Datav avaamista voidaan tarkastella prosessina, joka muodostuu useista vaiheista. Monet kuvatuista vaiheista voidaan toteuttaa itsenäisinä ja irrallisina toimenpiteinä. Kun päämääränä pidetään avoimen datan tuottamista, on tärkeää kuitenkin huolehtia siitä, että eri toimenpiteet tukevat toisiaan ja että ne tehdään

---

<sup>1</sup> <http://www.merikotka.fi/digiport>

keskenään yhteensopivasti. Tässä prosessissa satamanpitäjän organisaation tietoarkkitehtuurilla ja määrittelyillä toimintatavoilla on lopputuloksen kannalta merkittävä rooli. Kerätystä datasta ja sen luonteesta (staat- tinen tai dynaaminen) riippuen, avaaminen voi olla joko kertaluonteinen toiminto tai jatkuva prosessi. Myös staattisen datan ajantasaisuus on tärkeä tarkistaa säännöllisesti.

Satamissa **staattisiin** eli liikennejärjestelmää ja väyläsuunnittelua sekä kunnossapidon ohjelmointia ja seu- rantaa tukeviin tietovarantoihin kuuluvat väyläverkon, taitorakenteiden<sup>2</sup> sekä varusteiden ja laitteiden ra- kenne- ja kuntotiedot sekä kunnossapidon toimenpidetiedot. **Dynaamisiin**, liikennetilannetta kuvaaviin tie- toaineistoihin kuuluvat puolestaan tie- ja rataverkon ajantasaiset liikenne- ja häiriötiedot, alusliikenteen si- jaintitiedot, reitit, satamakäynnit ja aikataulut, ratatöiden ja junaliikenteen yhteensovittamisen ennakkotie- dot sekä väyläverkon olosuhteita kuvaavat sää- ja kelitiedot. Meriliikenteen tietovaranto sisältää myös alus- rekisteritiedot ja alusten ominaisuustiedot sekä Suomen vesialueet kattavat merenmittaustiedot.<sup>3</sup>

Mikäli avattava data on eri sijainneissa (eri tietokannat, palvelimet tms.) tai erilaisissa formaateissa, voidaan avaamisprosessi jakaa useampiin projekteihin datan sijainnin, tietotyypin tai -lajien perusteella ja toteuttaa jatkuvana prosessina. Oppaassa Julkinen data - johdatus tietovarantojen avaamiseen<sup>4</sup>, datan avaamisen vai- heet määritellään seuraavasti: datan analysointi, julkaiseminen ja prosessista oppiminen.

Kuvassa 1 datan avaamisprosessi on kuvattu ympyränä. Avaamisprosessissa eri vaiheita voidaan kytkeä yh- teen tai jakaa pienempiin osiin. Tämä riippuu kulloinkin avattavasta datasta, sen tietomuodosta ja käytössä olevista resursseista.



Kuva 1. Datan avaamisprosessi ympyräkuvaajana.

<sup>2</sup> rakenteet, joiden rakentamiseksi on laadittava lujuuslaskelmiin perustuvat suunnitelmat, tyypillisesti esimerkiksi sillat

<sup>3</sup> HE 61/2018 ([https://www.eduskunta.fi/FI/vaski/HallituksenEsitys/Documents/HE\\_61+2018.pdf](https://www.eduskunta.fi/FI/vaski/HallituksenEsitys/Documents/HE_61+2018.pdf))

<sup>4</sup> Antti Poikola, Petri Kola, Kari A. Hintikka, LVM 2010

Prosessi voidaan aloittaa useammasta kohdasta ympyrää, mutta usein datan avaamisen toteutus etenee seuraavassa järjestyksessä.

**Datan kartoitus;** satamanpitäjän organisaation luomat, keräämät, tuottamat tai hallinnassa olevat data-aineistot inventoidaan ja listataan jatkotoimenpiteitä varten. Avattavaksi suunniteltujen datasettien valinta on ensimmäisiä askeleita tässä prosessissa. Koko avaamisprosessia voidaan kuvata jatkuvaksi ja on aina mahdollista, ja jopa suositeltavaa, palata datan kartoitusvaiheeseen uudelleen. Datan kartoituksen lisäarvona on kokonaiskuvan muodostuminen tai sen paraneminen organisaation hallussa olevista datavarannoista. Monesti organisaatiolla ei ole tarkkaa kokonaiskuvaa siitä, mitä tietoja sillä on hallussaan. Tämän seurauksena organisaation kyky ja valmius tarjota dataa suoraan hyödynnettäväksi sekä lisäarvopalvelujen raaka-aineeksi saattaa jäädä vähäisemmäksi, kuin sen todellinen datavaranto mahdollistaisi.

**Datan luokittelu;** data voidaan luokitella eri ryhmiin<sup>5</sup>:

- julkinen (aiemmin jaettu)
- saatavilla (data on saatavilla, muttei vielä avattu tai käyttöä on osittain rajoitettu)
- avattavissa (data, joka voidaan avata, mutta teknisistä tai tekijänoikeudellisista syistä on toistaiseksi avaamatta), kun estävät tekijät on ratkaistu
- ei-julkinen (salaiseksi luokiteltu tai sensitiivinen informaatio, joka esimerkiksi muuhun tietoon yhdistettynä muodostaa yksityisyydensuoja-, turva- tai muun riskin)

Samassa yhteydessä voidaan selvittää ja määritellä dataan mahdollisesti liittyvät tekijänoikeusseikat sekä sovitavat lisensiointimallit julkaisua varten.

**Käyttäjätarpeet (vaatimukset);** tehdään arvio ja selvitys oletettujen käyttäjäryhmien tarpeista ja vaatimuksista. Tämä vaihe on paras toteuttaa yhteistyössä ajateltujen käyttäjäryhmien kanssa. Näitä voivat olla esimerkiksi satamassa toimivat ja siihen liittyvät yritykset, huollosta- ja kunnossapidosta vastaavat toimijat sekä dataa hyödyntävät lisäarvopalvelujen tuottajat. Samassa yhteydessä voidaan datan edelleen jalostavien palveluntuottajien kanssa käydä keskustelua soveltuvimmista tietomuodoista ja esitystavoista.

**Tietomuotojen määrittäminen;** yleensä tietokantaan kerätyn datan tietomuotoa ei vaihdeta julkaisua varten, etenkin jos tieto on standardoidussa muodossa. Erityisesti tässä vaiheessa standardinmukaisten tietomuotojen ja harmonisoitujen tietosisältöjen edut tulevat selkeästi esiin työn ja kustannusten säästönä.

**Julkaisu;** on datan avaamisen tärkeimpiä vaiheita. Avoimesta datasta on kovin vähän hyötyä, jos sen potentiaaliset käyttäjäryhmät eivät tiedä sen olemassaolosta ja saatavuudesta. Siksi on tärkeää **tiedottaa avatusta datasta** mahdollisimman laajalti potentiaalisille käyttäjille ja käyttäjäryhmille. On myös tärkeää julkaista data mahdollisimman helposti saatavana ja tiedottaa rajapinnasta, jonka kautta data on ladattavissa.

**Vuorovaikutus käyttäjien kanssa ja palautteen kerääminen;** yhteydenpitoa käyttäjäryhmien kanssa ja käyttäjäpalautteen keräämistä tulisi jatkaa datan avaamisen ja julkaisun jälkeen. Käyttäjäpalautetta voidaan kerätä avatusta tietosisällöstä, käyttökelpoisuudesta, kattavuudesta ja täsmällisyydestä sekä käyttökokemuksista, lataamisen helppoudesta ja oikean tiedon löydettävyydestä. Samoin myös kokemusten vertailu muiden satamien kanssa datan avaamisesta, on erinomainen keino parantaa käytettävyyttä. Käyttäjäpalautteen keräämisratkaisuina voidaan käyttää esimerkiksi suoria kyselyjä ja keskusteluja sekä erilaisia sähköisiä verkkosivuja sekä keskustelualustoja.

Kun datan avaamisprosessin ensimmäinen kierros on tehty, olisi hyvä kerätä myös sisäisen prosessin aikana saadut kokemukset ja palaute sekä hyödyntää niitä seuraavilla avaamiskierroksilla tekemällä tarvittavat muutokset ja parannukset toimintatapaan ja avattavan datan käsittelyyn.

---

<sup>5</sup> Luokittelu ei vastaa asiakirjojen luottamuksellisuudesta käytettävää luokitusta

### 3.2. Datan yhdenmukaistaminen ja standardisointi

Samansisältöistä dataa saatetaan erilaisista syistä (esim. sen hetkinen käytäntö, tarve tai käyttötarkoitus) kerätä ja tallentaa eri organisaatioissa ja ajankohtina erilaisissa muodoissa. Tällöin tietokantoihin ja järjestelmiin kertyy dataa, jossa samasta tiedosta ei käytetä samaa nimeämistapaa. Tämän vuoksi datan avaamisen yhteydessä on hyvä selvittää datan yhdenmukaistamis- ja standardisointimahdollisuudet.

Datan yksinkertaistamis- ja yhdenmukaistamisprosessi koostuu neljästä vaiheesta: **datan kerääminen, määrittely, analysointi ja yhdenmukaistaminen**<sup>6</sup>. Jotta datan avaamisesta saataisiin irti paras mahdollinen hyöty, yhdenmukaistaminen tulisi tehdä jo datan keräämisvaiheessa ja viimeistään ennen datan avaamista. Datan yhdenmukaisen nimeämisen lisäksi, on suositeltavaa käyttää datan standardinmukaista esitystapaa, kuten YK:n tai ISO:n<sup>7</sup> koodistoja.

Datan yksinkertaistamisen ja yhdenmukaistamisen tavoitteena on vähentää toiminnan tietovaatimuksia ja selkeyttää tietorakenteita poistamalla ja yhdistämällä päällekkäisiä ja samansisältöisiä tietoelementtejä. Tämän prosessin lopputuloksena saadaan tehokkaampi datan hyödyntäminen tietoa tuottavien, toimittavien ja keräävien osapuolten kesken.

Satamien avatessa (ja yhdenmukaistaessa) dataa, olisi suositeltavaa, mahdollisuuksien mukaan, noudattaa samaa nimeämiskäytäntöä, kuin keskeiset datarekisterejä ylläpitävät liikenneviranomaiset.

Datan yksinkertaistaminen ja yhdenmukaistaminen on vaativa ja resursseja kuluttava tehtävä. Se edellyttää sekä sataman toimintojen että niissä tarvittavien tietojen käytön hyvää tuntemusta. Silti yhdenmukaistaminen kannattaa pidemmällä tähtäimellä, koska se helpottaa ja tehostaa merkittävästi datan jatkokäyttöä ja tarjottavien palvelujen tuottamista, sekä sataman omissa toiminnoissa, että ulkopuolisten palvelutoimittajien palveluissa.

## 4. Sataman avoin data ja tiedonhallinta

Avoimen datan hallintaan tulee kiinnittää huomiota, jotta voidaan varmistua datan ja siihen kohdistuvien toimenpiteiden yhdenmukaisuudesta sataman muun tiedonhallinnan kanssa, käytön tehokkuudesta ja helpoudesta. Datan suojaamiseen ja luotettavuuteen tulee niin ikään kiinnittää huomiota, jotta asiaan kuulumattomat tahot eivät pääse esimerkiksi muuttamaan tai tuhoamaan sitä. Siksi onkin hyvä nimetä satamanpitäjän jaettavasta tiedosta vastaava henkilö.

Tiedonhallinnan tukena ja tietojen suojaamisen apuna voidaan käyttää esimerkiksi julkisen hallinnon digitaalisen turvallisuuden johtoryhmän (VAHTI) ohjeita, erityisesti GDPR:n mukaisesti uudistunutta tietosuojalakia ja muita asiaan liittyviä julkisen hallinnon suosituksia (JHS)<sup>8</sup>.

Avoimen datan käyttäjille ja varsinkin sitä hyödyntävien palvelujen tarjoajille on erityisen tärkeää, että avoin tieto, jota he käyttävät, on luotettavaa. Samoin, jos datassa on puutteita tai muita poikkeavuuksia, käyttäjät saavat siitä tiedon.

Tiedonhallinnan tueksi ja toimenpiteiksi satamanpitäjien organisaatioissa ehdotetaankin DigiPort-hankkeessa, että:

---

<sup>6</sup> [https://www.unece.org/fileadmin/DAM/trade/Publications/ECE-TRADE-400E\\_Rec34.pdf](https://www.unece.org/fileadmin/DAM/trade/Publications/ECE-TRADE-400E_Rec34.pdf)

<sup>7</sup> International Organization for Standardization

<sup>8</sup> Tietosuojalaki (2018), JHS-suositukset, julkisen hallinnon digitaalisen turvallisuuden johtoryhmän (VAHTI) ohjesivusto [www.vm.fi/vahti](http://www.vm.fi/vahti).

- datan kartoitustyö ja avaamisprosessi voidaan kuvata osana sataman laatukäsikirjaa. Samoin tehdään myös avoimeen dataan liittyvien tiedonhallinnan toimenpiteiden ja vastuiden osalta, etenkin silloin, jos ne poikkeavat satamanpitäjän normaaleista tietohallintorutiineista.
- sataman laatukäsikirjaan työ- tai/ja menettelyohjeisiin dokumentoidaan datan avaamiseen liittyvät keskeisimmät toimenpiteet
- tarkistetaan laatukäsikirjan politiikkaa: miten avoin data on tuotu siinä esille?
- laatukäsikirjaan liitetään johdon katselmuksen osaksi myös avointen data-aineistojen julkaisemisen hyväksymismenettely
- laatukäsikirjaan laaditaan avoimen datan laatumittarit ja poikkeamailmoitusmenettely datan käyttäjille ja hyödyntäjille.

Mikäli datan kartoituksessa ja luokittelussa tullaan siihen lopputulokseen, että dataa tai joitakin sen osia ei jostain syystä (laki, säännökset, toimintapolitiikka, luottamuksellisuus, turvallisuus, kaupalliset tai sopimuskelliset syyt jne.) tällä hetkellä voida avata, tulisi päätös perusteluineen kirjata esimerkiksi laatukäsikirjaan. On suositeltavaa, että satamanpitäjä määrittelee etukäteen kriteeristön, joka rajoittaa tai estää datan avaamisen. Rajoittavia kriteerejä voidaan datan kartoitustyössä, vielä hankkeen aikanaan, arvioida ja testata tai viimeistään katselmuskokouksessa, jossa avattava data hyväksytään. Jos katselmuksessa löytyy datasettejä, jotka läpäisevät kriteeristön, mutta joita ei haluta avata, niin siltä osin vastuuhenkilö, esimerkiksi laatupäällikkö, täydentää kriteeristöä. Tällä tavoin päätös on dokumentoitu ja tilanteen mahdollisesti muuttuessa päätöstä voidaan tarkastella kaikki siihen vaikuttaneet seikat huomioon ottaen.

Datan kartoituksen ja luokittelun yhteydessä luodaan myös lista, josta ilmenevät ne tiedot, jotka on avattu tai voidaan avata. Listan avulla yleisö ja ao. datasta kiinnostuneet tahot voivat seurata avaamisprosessin edistymistä.

## 5. Standardit, datan esitystavat ja rakenteet

Avoimen datan ja standardien käytön välillä on positiivinen vaikutussuhde. Mitä laajemmin dataa jaetaan ja hyödynnetään, sitä enemmän se tuottaa lisäarvoa. On selvää, että standardeihin pohjautuva tietosisältö ja datan esitystapa mahdollistavat laajemman datan hyödyntämisen, rikastamisen ja jalostamisen kuin järjestelmä- tai sovelluskohtaisessa (esim. itse määritellyt koodit tai ilmaisut) muodossa avattava sisältö tai esitystapa.

Siksi satamia, jotka keräävät ja avaavat dataa, suositellaan tallentamaan ja avaamaan datansa standardin mukaisessa muodossa (sekä sisällön että esitystavan osalta). On huomattava, että datan sisältöstandardi (esim. koodisto) ja esitystapa ovat eri standardeja, esimerkkinä satamien tunnisteet (UN/LOCODE)<sup>9</sup> ja esitystapastandardi (JSON<sup>10</sup> tai CSV<sup>11</sup>).

Datan esittämiseen on olemassa useita esitystapastandardeja eri tarkoituksiin. Pääsääntöisesti käytettävän standardin valinta on selkeää ja suoraviivaista ja lähtee sataman tarpeista. Voi kuitenkin olla tilanteita, joissa eri satamanpitäjät ovat toiminnallisista tai muista syistä valinneet samanlaisen datan tallentamiseen eri sisältö- tai esitystapastandardeja. Ohjeena standardin valinnassa voidaan pitää valtakunnan tasolla toiminnasta vastaavan ja datavarantoa ylläpitävän viranomaisen (esim. liikenneviranomaisen) kyseisestä tiedosta käyttämää sisältö- ja esitystapastandardia. Mikäli taas useat toimijat kokoavat ja tallentavat itsenäisesti samaa tai lähes samaa tietoa eri tarkoituksiin, on erityisen suositeltavaa pyrkiä toteuttamaan tietosisältöjen ja esitystapojen yksinkertaistamis- ja yhdenmukaistamisprosessi.

<sup>9</sup> [UNLOCODE](#)

<sup>10</sup> [ECMA-404](#)

<sup>11</sup> [CSV](#)

Liikenteen hallinnassa ja siihen liittyvän datan esitystapana käytetään varsin yleisesti JSON, CSV ja XML -formaatteja.

## 5.1. Lyhyt katsaus eräisiin avoimen datan esitystapoihin

**JSON** (JavaScript Object Notation)<sup>12</sup> on JavaScript-ohjelmointikielen osajoukko ja yksinkertainen tiedostomuoto, jonka prosessointi on helppoa erilaisilla ohjelmointikielillä. Yksinkertaisuus tarkoittaa sitä, että vaikka JSON:in esitysvoima on rajoitettu, sen prosessointi on helppoa ja kevyttä esimerkiksi SGML:n verrattuna. JSON:in tekninen määrittely on ECMA-404 -standardissa<sup>13</sup>.

JSON on kevyt myös tiedonsiirtomuotona. Ihmisten on helppo lukea ja kirjoittaa sitä, sen prosessointi ja jäsenitys on myös helppoa. Se on tekstimuotoista ja täysin kieliriippumatonta, mutta tukee ohjelmoijien yleisesti käyttämiä rutiineja. JSON voi hyödyntää Table Schemaa<sup>14</sup> sarakemuotoisen rakenteen tuottamiseksi.

**CSV** (Comma Separated Value) on varsin käyttökelpoinen esitystapa, koska se on rakenteeltaan tiivis ja soveltuu taulukkomuotoisen datan esittämiseen tekstinä, sekä suurten, samanrakenteisten tietojoukkojen siirtämiseen. CSV-esitystapa on kuitenkin niin riisuttu, että vaikka tietosisältö olisi helposti ihmisen luettavissa, sen tulkinta ja hyödyntäminen voi olla jopa mahdotonta ilman siihen liittyvää dokumentaatiota (metadata). Siksi on tärkeää, että tekstisarakkeiden sisältö on tarkasti kuvattu ja määritelty. Tämä voidaan tehdä joko tiedoston alussa (ensimmäisellä rivillä) tai erillisenä tiedostona.

Lisäksi on tärkeää, että määriteltyä rakennetta noudatetaan tarkasti ja esimerkiksi tyhjät sarakkeet, eli puuttuva tieto, esitetään erotinmerkeillä. Muuten seuraavien tietojen luku, rakenne ja tulkinta häiriytyvät ja virhekohtaa on äärimmäisen vaikea havaita jälkeenpäin. Vaikka CSV on jo iäkäs esitystapa, sitä ei ole tarkasti standardoitu (RFC 4180), joten siksi UTF-8 merkistökoodauksen käyttö CSV:n yhteydessä on suositeltavaa.

**XML** (Extensible Markup Language) on laajalti käytetty esitystapa, koska se tarjoaa mahdollisuuden säilyttää datan ja tiedostojen rakenteen. Lisäksi se mahdollistaa selkeyttävän dokumentaation ja merkinnät datatiedostossa – datan lukemisen ja prosessoinnin siitä kärsimättä.

**RDF** (Resource Description Framework) on World Wide Web Consortiumin (W3C) standardoima esitystapa metatiedon kuvaukseen, jonka avulla on mahdollista yhdistää tietoja eri lähteistä. RDF-data voidaan tallentaa esimerkiksi JSON ja CSV-formaateissa. RDF:ssä voidaan käyttää tunnisteina [URL:ja](#), mikä mahdollistaa verkossa olevien avoimen datan lähteiden suoran kytkennän. RDF ei ole vielä kovin yleisesti käytössä, mutta sitä on käytetty avoimen julkishallinnon hankkeissa.

**Laskentataulukko** on erittäin yleisesti käytössä esitettäessä ja julkaistaessa dataa, erityisesti luku- ja aikasarjoja, mutta myös muissa datajoukoissa. Esimerkki laskentataulukosta on Microsoft Excel. Tällaista dataa on usein mahdollista käyttää sellaisenaan, kun riittävä metadata eli kuvaukset sarakkeiden sisällöstä ja merkityksestä on annettu.

Laskentataulukoissa saattaa joskus olla makroja ja laskentakaavoja, jotka hankaloittavat taulukon käyttöä ja käsittelyä. Siksi onkin suositeltavaa dokumentoida makrojen ja laskentakaavojen avulla tuotetut laskelmat taulukon viereen, koska ne ovat silloin helpommin käyttäjän saatavilla.

**Tekstiasiakirja** on formaattina (kuten Word, ODF, OOXML tai PDF) käyttökelpoinen joillekin tietotyypeille, kuten stabiileille osoitetiedoille ja vastaaville. Normaalisti tämä on kyseisen datan alkuperäinen esitystapa ja

---

<sup>12</sup> Lisätietoja JSON/JSONP: <http://en.wikipedia.org/wiki/JSON> ja <http://json.org/>

<sup>13</sup> [RFC 4627](#) and ECMA-404 standard <http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf>

<sup>14</sup> <https://frictionlessdata.io/specs/table-schema/>



siten helppo julkaista. Esitystapa ei kuitenkaan tue datan yhtenäistä rakennetta ja sen säilymistä, mikä tarkoittaa usein hankaluuksia data automaattiselle lisäämiselle ja käsittelylle. Yleisesti ei olekaan suositeltavaa esittää dataa tekstiasiakirjan muodossa, jos data on saatavilla muissa formaateissa.

**Tekstitiedosto** (.txt) on helppolukuinen myös tietokoneelle. Tietokone jättää yleensä käsittelemättä rakenteisen metadatan dokumentin sisällä, mikä merkitsee sitä, että ohjelmoijien pitää laatia jäsentämiskäsitelmä, jonka avulla dokumentti voidaan tulkita julkaisumuodossaan. Ongelmia voi myös syntyä, mikäli joudutaan vaihtamaan tekstitiedostoja käyttöjärjestelmästä toiseen (Windows, MacOS, UNIX). Tekstidokumenteissa on suositeltavaa käyttää UTF-8 merkistöä.

## 5.2. Avoimen datan rajapinnoista ja saatavuudesta

Eräs tärkeimpiä tehtäviä datan avaamisen kokonaisuudessa on datan saatavuuden järjestäminen sen hyödyntäjille. Julkaisu- ja latausrajoitusten sekä niihin liittyvien menettelyjen selkeä määrittely ja toteutus ovat merkittäviä tekijöitä datan uudelleenkäytön helpoudessa.

Datan julkaiseminen ohjelmointirajapinnan (Application Programming Interface, API) kautta on hyvin suositeltavaa. Tämä antaa ohjelmoijille mahdollisuuden valita ladattavaksi haluamansa data sen sijaan, että koko datamassa olisi saatavilla vain yhtenä (suurena) tiedostona. Ohjelmointirajapinta on tyypillisesti kytketty tietokantaan, jota päivitetään jatkuvasti. API:n kautta tarjottava data on yleensä siis ajantasaisista.

Raakadata, joka on tietojen jalostuksen tai jonkin tietyn datan käsittelyvaiheen kannalta muokkaamatonta dataa, voidaan julkaista myös massana. Tällöin datan uudelleenkäytön ei tarvitse liittyä sen alkuperäiseen käyttötarkoitukseen. Raakadataa on myös helppo muokata ja konvertoida hyödyntäjän omiin tarpeisiin sopivaan muotoon, linkittää muuhun dataan ja resursseihin tai versioida ja arkistoida useisiin eri paikkoihin. Esimerkkinä raakadatasta on liikenteen automaattiselta mittausasemalta (LAM) saatu liikennedata.

Datan julkaisutavan valinnassa on hyvä huomioida datan tiedossa olevat käyttötavat ja valita julkaisumuoto siten, että se palvelee niitä – luonnollisesti avaamisen kustannukset huomioiden. Jos esimerkiksi ajantasainen data tarjotaan API:n välityksellä, voidaan raakadata tarjota bulkkina säännöllisin väliajoin päivitettyinä.

Yleinen tapa tällä hetkellä tarjota avointa dataa on perustaa tai tarjota suora yhteys datan avaajan ja hyödyntäjän välille. Tällöin datan avaaja toteuttaa rajapinnan tai tarjoaa tiedostot hyödyntäjien käytettäväksi. Hyödyntäjä voi tarpeidensa mukaan ja halutessaan ladata dataa rajapinnan kautta tai tiedosto(je)n muodossa.

DigiPort-hankkeessa perustetaan datakatalogi/portaali, joka tarjoaa erinomaisen mahdollisuuden avata sataman data hyödyntäjien saataville. Portaalin kautta tarjotulla datalla on monia hyötyjä sekä sovelluskehittäjille ja käyttäjille, että myös dataa avaaville satamille:

- Satamanpitäjä avaa rajapinnan ja sopii datan edelleen jakelusta vain yhden osapuolen (DigiPort-portaalin) kanssa
- Datan avaajan ja käyttäjän välillä ei ole välttämättä suoraa riippuvuussuhdetta.
- Kolmannet osapuolet (kuten aggregaattori) voi ladata kopion datasta, jalostaa sitä osana omaa palveluaan ja edelleen jakaa sitä. Tämä vähentää datan lähdesataman jakelukustannuksia, resurssitarvetta ja työmäärää.
- Jos tietojen saatavuudessa tai jakelussa on ongelmia, satamanpitäjän ei tarvitse selvittää niitä datan hyödyntäjien ja käyttäjien suuntaan – riittää, kun sataman ja portaalin välinen tiedonsiirto toimii.
- Kolmannet osapuolet voivat kehittää datapohjaisia palveluita vakuuttuneina siitä, että ladattu (bulkki) data on jatkuvasti heidän käytössään.

## 6. Lisensiointimallit avoimen datan tietokannoissa

Datan avaamisen ja siihen motivoinnin kannalta on tärkeää, että eri organisaatioiden avaamaa dataa myös käytetään mahdollisimman paljon ja sen avulla toteutetaan lisäarvoa tuottavia palveluja.

Avoim data saattaa olla tekijänoikeuden, lähioikeuden tai luettelosuojan alaista. Avoimen datan lisensoinnilla annetaan oikeus käyttää dataa omaan tarkoitukseen ja levittää sitä eteenpäin. Avoimen datan käyttö ja siitä palvelujen tuottaminen on organisaatiolle varsin edullista; data on ilmaista, organisaation ei tarvitse käydä neuvotteluja datan käytöstä eikä maksaa juristipalkkioita sopimusten laatimisesta.

Avoimen data lisensointimalleja tuottaa Creative Commons<sup>15</sup> (CC) -yhteisö. Se on laatinut eri tarkoituksiin erilaisia lisensointimalleja, jotka ovat modulaarisia, eli niissä voidaan viitata eri maiden lainsäädäntöön esim. riitojenratkaisussa, muuttamatta lisensointiluvan muita ehtoja.

Suomessa myös julkishallinto on laatinut suosituksen avoimen datan lisensointimalleista (JHS 189)<sup>16</sup> tätä suositusta voidaan seurata ja soveltaa myös satamien avoimen datan lisensoinnissa.

Julkishallinnon suositus 189 suosittelee Creative Commons Nimeä 4.0 -lisenssin käyttämistä julkishallinnon tietoaineistojen julkaisemiselle, prosessoinnille ja uudelleenkäyttämiseksi. Kyseinen lisenssi antaa oikeudellisen varmuuden siitä, että käyttöluvan kohteena olevaa dataa voi vapaasti käyttää kaikin mahdollisin tavoin edellyttäen, että datan lähde mainitaan käyttöluvasta ilmenevällä tavalla. Käyttöoikeuksien tulee olla kone-luettavassa muodossa hyödyntämisen helpottamiseksi ja niiden tulisi olla teknisesti yhteensopivia muiden avoimen datan lisenssien kanssa.

### 6.1. Creative Commons -lisenssit

Creative Commons -ehdot ovat maailmanlaajuisesti useimmin käytettyjä avoimen datan lisensointiehtoja. Lisenssien periaatteena on, että materiaalit ovat vapaassa käytössä edellyttäen, että käyttöehtoja noudatetaan. Yleisin ehto on lisensoijan nimeäminen (CC BY).

Avoimen datan lisenssit on pyritty laatimaan selkeiksi ja yksinkertaisiksi ja niitä on myös selvästi vähemmän kuin esimerkiksi avoimen lähdekoodin lisenssejä. Koska avoimen datan Creative Commons -lisenssit ovat modulaarisia ja samansisältöisiä, ne voidaan myös esittää yksinkertaisesti ikonin (kts. kappale 6.4) avulla.

#### 6.1.1. Creative Commons nimeä 4.0

Creative Commons nimeä 4.0 -lisenssi vastaa parhaiten niitä periaatteita ja vaatimuksia, jotka on asetettu esimerkiksi julkishallinnon avoimelle datalle. Se sisältää myös tietokantojen tekijänoikeuksiin soveltuvat ”sui generis” -periaatteet, joita sovelletaan EU:n jäsenvaltioissa ja Euroopan talousalueen maissa.

Creative Commons nimeä 4.0 on peruuttamaton julkinen lisenssi, joka sallii tekijänoikeuden ja muiden vastaavien lähioikeuksien suojaamien teosten ja muun materiaalin jakamisen korvauksetta yleisölle määriteltujen vakioehtojen mukaisesti. Julkaistun avoimen datan lisenssin ehtoja ei voi jälkikäteen muuttaa takautuvasti.

Creative Commons nimeä 4.0 ei anna yksinomaista käyttöoikeutta lisensoituun dataan ja tietokantoihin. Lisensiointi koskee datan tekijänoikeuksia ja lähioikeuksia, ja siksi lisensoijalla pitää olla valtuudet myöntää

---

<sup>15</sup> <https://creativecommons.fi/lisenssit/>

<sup>16</sup> <http://www.jhs-suositukset.fi/suomi/jhs189>

käyttölupa kyseiselle datalle. Tämä merkitsee muun muassa sitä, että julkaisijalla tulee olla oikeus julkaista kyseinen data ja julkaisija on hankkinut oikeudet kolmannen osapuolen dataan (esim. avoimen datan lisenssi), joka mahdollisesti on sisällytetty julkaistavaan dataan. Kolmannen osapuolen dataa hankittaessa olisi hankintasopimuksessa varauduttava tarvittaessa siihen, että kyseinen data tai osia siitä, voidaan julkaista esimerkiksi muun avattavan datan osana.

## 6.2. Julkaistava data ja siihen liittyvät rajoitukset

Julkishallinnon asiakirjat (mukaan lukien sähköisessä muodossa oleva ja vain tietojenkäsittelyn avulla saatavissa oleva data-aineisto) ovat laajalti julkisia. Viranomaistietojen julkaisua sääntelee Julkisuuslaki. Yksityiset organisaatiot voivat periaatteessa vapaasti päättää, mitä omistamiaan data-aineistoja ne julkaisevat.

Julkisuuslaissa on säädetty yleisistä asiakirjojen julkisuusperiaatetta rajoittavista salassapitoperusteista. Julkisuuslain 24§:n mukaan salassa pidettäviä asiakirjoja ovat muun muassa valtion yleiseen ulkoiseen ja sisäiseen turvallisuuteen liittyvät asiakirjat, uhanalaisiin eläin- ja kasvilajeihin liittyvät asiakirjat, yksityisestä liike- tai ammattisalaisuudesta tietoja antavat asiakirjat sekä tilastoviranomaiselle tilastojen laatimista varten annetut asiakirjat mukaan luettuna asiakirjat, jotka on vapaaehtoisesti annettu viranomaiselle tutkimusta tai tilastointia varten.

Säännös soveltuu luonnollisesti myös yksityisen organisaation toimintaan ja se tarkoittaa käytännössä, että tämä yleistä käyttölupaa koskeva suositus ei koske tietoaineistoja, jotka julkisuuslaissa on määritelty salassa pidettäviksi. Myös yksityisyyden suojaava henkilötietolaki ja tekijänoikeuslaki asettavat aineiston uudelleenkäytölle rajoituksia.

Tämän vuoksi Creative Commons nimeä 4.0 -käyttölupaa ei sovelleta aineistoihin, joihin sisältyy kolmansien tahojen immateriaalioikeuksia, ellei näitä oikeuksia ole hankittu käyttöluvan myöntävälle organisaatiolle. Koska käyttölupa koskee vain tekijänoikeuteen perustuvia käyttöoikeuksia, sillä ei voi myöntää käyttölupaa henkilötietoihin, joiden käsittelyä ja luovutusta on rajoitettu muussa lainsäädännössä.

## 6.3. Tietokantojen tekijänoikeudet

Tietokannan tekijänoikeudellinen suoja koskee yleensä vain tietokannan rakennetta, sitä miten tietokanta ja tiedot on järjestetty sekä menetelmää, jolla tietokannassa olevat tiedot on kerätty. Tietokantoja voi tekijänoikeudellisesta näkökulmasta olla kolmenlaisia:

1. kirjallisena teoksena suojatut tietokannat
2. sui generis -oikeudella suojatut tietokannat
3. luettelot.

### 6.3.1. Kirjallisena teoksena suojatut tietokannat

Tietokannan, joka suojataan kirjallisena teoksena, tulee osoittaa tekijänsä omaperäistä ja luovaa panosta. Tietokannan tekijällä on tällöin lainmukainen oikeus määrätä tietokannan kappaleen valmistamisesta ja saattamisesta yleisön saataviin. Kirjallisena teoksena suojattava tietokanta nauttii maailmanlaajuista suojaa Bernin sopimuksen ja WTO:n TRIPS-sopimuksen pohjalta.

### 6.3.2. Sui generis- ja luettelosuojalla suojatut tietokannat

Sellaisia tietokantoja, jotka eivät täytä tekijänoikeudelliselle suojalle asetettuja vaatimuksia suojataan EU:n alueella niin sanotulla sui generis -tietokantaoikeudella. Sui generis -oikeudella suojattavien tietokantojen sisällön kerääminen, todentaminen tai esittäminen on edellyttänyt huomattavaa panostusta. Suoja-aika on

15 vuotta tietokannan valmistumisesta ja se jatkuu aina uuden 15 vuoden jakson, kun tietokantaa oleellisesti muutetaan.

Tietokannat, jotka eivät saa tekijänoikeus- tai sui generis-suojaa, voivat Suomen tekijänoikeuslain mukaan olla suojattuja luettelosuojan perusteella, mikäli ne sisältävät suuren määrän tietoja.

Luettelosuoja on perinteinen pohjoismaisessa tekijänoikeuslainsäädännössä oleva suojamuoto, jonka tarkoituksena oli suojata puhelinluetteloiden ja muiden vastaavien luetteloiden valmistajien työpanosta. Luettelosuoja koskee luetteloita, joihin on kerätty suuri määrä tietoja ja se on voimassa 15 vuotta luettelon valmistamisesta.

Erityisesti sui generis- ja luettelosuojan erottaminen toisistaan, voi olla vaikea yksittäisissä tapauksissa. Julkishallinnon suosituksessa 189 suositellaan, että lupa kaikenlaisten tietokantojen käyttöön myönnetään aina Creative Commons nimeä 4.0 -lisenssillä.

### 6.3.3. Raakadataa sisältävät tietokannat

Raakadataa sisältävät tietokannat ovat tyypillisesti sellaisia, joihin kerätään automaattisen mittauslaitteiston avulla esimerkiksi liikennetietoja (LAM-mittauspisteet), ilmakehän kaasuja, säätietoja ja vastaavaa luonnon-ilmioita koskevaa dataa.

Tallennettu tai mitattu data sinänsä, eli raakadata, ei ole suojattu tietokantana, mutta siihen voi tietyin edellytyksin kohdistua sui generis -tietokantaoikeus. Sui generis -suojan edellytyksenä on, että tietokannan sisällön kerääminen, varmistaminen tai esittäminen merkitsee määrällisesti ja/tai laadullisesti huomattavaa investointia.


Datan käyttäjälle on aina turvallisinta ilmaista datan lähde, kuin ottaa riski siitä, että lähdemateriaali on tekijänoikeudella tai muilla ehdoilla suojattua tekijän kotimaassa.


Jos esimerkiksi laaditaan sovellus, jossa visualisoinnin avulla kuvataan sataman liikennemääriä ja -virtoja, pohjautuen kerättyyn raakadataan, sovelluksen laatijoilla on tekijänoikeus sovellukseen ja he voivat lisensoida sen haluamallaan tavalla. Sovelluksen tulee kuitenkin sisältää tieto lähdedatasta sekä linkki alkuperäiseen dataan, joka on lisensoitu Creative Commons nimeä 4.0 -ehdoilla.

## 6.4. Lisensointimalleista

Jotta avoin data olisi käyttökelpoista eri osapuolille, kuten esimerkiksi sovellusten kehittäjille, avoin data on syytä julkaista selkeillä lisensiointiehdoilla varustettuna. Tämä on oleellista myös epäselvyyksien välttämiseksi dataa uudelleen käytettäessä ja hyödynnettäessä.

Datan julkaisemista varten on olemassa useita erilaisia lisensoijia rajoitetuista täysin avoimiin. Näitä esittelee taulukko 1. Sarakkeella ”domain” tarkoitetaan toimialuetta, eli minkä tyyppiseen materiaaliin (dataan) lisenssiä voidaan käyttää. Seuraavassa yleisimmin käytetyt ja suositellut Creative Commons -lausekkeet ja niitä kuvaavat ikonit:

- BY = edellyttää (lähteen) nimeämistä  <sup>17</sup>

- SA = edellyttää edelleen jakamista samalla lisenssillä.  <sup>18</sup>

<sup>17</sup> <https://creativecommons.org/>

<sup>18</sup> <https://creativecommons.org/>

### 6.4.1. Suositeltavia lisensointimalleja

Alla olevat lisenssimallit täyttävät avoimen tiedon määritelmän periaatteet<sup>19</sup>:

- Uudelleen käytettäviä: ei yritys- tai lainsäädäntökohtaisia.
- Yhteentoimivia: Pitää olla yhteensopiva GPL-3.0+, CC-BY-SA-4.0, ja ODbL-1.0 kanssa ja vähintään yhden seuraavista: Apache-2.0, CC-BY-4.0, ja ODC-BY-1.0 kanssa.
- Ajankohtaisia: laajalti käytettyjä ja soveltamisalueellaan laajan toimijakunnan yleisesti hyväksi käytännöiksi tunnustamia lisenssejä.

Taulukko 1. Lisensointimalleja.

Lisenssi	Domain	BY	SA	Kommentit
<a href="#">Creative Commons CCZero</a> (CC0)	Content, Data	N	N	Dedicate to the Public Domain (all rights waived)
<a href="#">Open Data Commons Public Domain Dedication and Licence</a> (PDDL)	Data	N	N	Dedicate to the Public Domain (all rights waived)
<a href="#">Creative Commons Attribution 4.0</a> (CC-BY-4.0)	Content, Data	Y	N	
<a href="#">Open Data Commons Attribution License</a> (ODC-BY)	Data	Y	N	Attribution for data(bases)
<a href="#">Creative Commons Attribution Share-Alike 4.0</a> (CC-BY-SA-4.0)	Content, Data	Y	Y	
<a href="#">Open Data Commons Open Database License</a> (ODbL)	Data	Y	Y	Attribution-ShareAlike for data(bases)

Creative Commons on laatinut verkkosivun: <https://creativecommons.org/choose/>, jonka avulla sopivan lisenssin valinta on pyritty tekemään mahdollisimman yksinkertaiseksi.

## 7. Avoimen datan varantojen ylläpidossa huomioitavia asioita

### 7.1. Datun laadun ylläpito

Hyvien käytäntöjen ja standardien noudattaminen datan keräämisessä ja tallentamisessa helpottaa datan avaamista ja avoimeen dataan kohdistuvien laatuvaatimusten täyttymistä. Lähtödatan laadusta huolehtiminen jo sen alkulähteellä voi vähentää merkittävästi datan julkaisukuntoon jalostamista ja sen julkaisemiseen liittyvää valmistelutyötä. Tässä olisi hyvä huomioida mahdollisuuksien mukaan myös datan käyttäjien toiveet ja näkemykset.

Data tulisi julkaista sen sisältöä ja laatuominaisuuksia kuvaavalla tiedolla (metadata) varustettuna. Tämä auttaa datan mahdollisia käyttäjiä hahmottamaan, onko data heidän tarpeisiinsa sopivaa ja keskittämään selvitystyönsä ja käyttötestauksensa oikeanlaiseen dataan. Esimerkiksi tilastollisen datan laatua arvioitaessa voidaan tunnistaa seitsemän tekijää: saatavuus, täsmällisyys, ajankohtaisuus, yhtenäisyys, yhteentoimivuus, merkityksellisyys ja konteksti.

<sup>19</sup> <https://opendefinition.org/od/1.1/fi/>

Yhdessä metadatan kanssa laatuominaisuuksien kuvaus antaa varsin hyvän käsityksen datan hyödyntäjille tietojoukkojen käyttökelpoisuudesta heidän tarpeisiinsa. Datan laatukysymykset eivät kuitenkaan saisi hidastaa tarpeettomasti tietojoukkojen julkaisemista.

## 7.2. Datan yhtenäisyyden varmistaminen

Avoimen datan ja sen lähteen luotettavuus sekä avattavan datan yhtenäisyys ovat erittäin tärkeitä datan hyödyntäjille ja tiedon jalostajille. Näiden asioiden tulee olla kunnossa, jotta kyseisen datan pohjalle on mahdollista perustaa palvelutuotantoa.

Data olisi periaatteessa hyvä julkaista sellaisena, kuin se on tiedon lähteeltä kerätty, kaikkine tietosisältöineen ja mahdollisimman vähän yhdisteltynä tai muokattuna. Yhdistelyn ja muokkauksen voivat tehdä esimerkiksi palveluntuottajat ja datan loppukäyttäjät omiin tarpeisiinsa sopivimmalla tavalla. Toisaalta tiedon jatkojalostajat haluaisivat datan mahdollisimman valmiina ja tehdä niin vähän virheriskiä lisääviä ja (koodaus)työtä vaativia muunnoksia kuin mahdollista.

Avoimen datan käyttömahdollisuudet ovat mittavat sekä hyödyntämisalueet ja -tavat varsin laajat. Siksi myös asiat, jotka liittyvät datan luotettavuuteen ja yhtenäisyyteen, ovat moninaiset ja datan julkaisumuoto tai -muodot on valittava harkiten.

Seuraavassa käsitellään muutamia avoimen datan ja sen lähteiden luotettavuuteen ja yhtenäisyyteen liittyviä aiheita:

**Luotettavuus:** Satamanpitäjän data pohjautuu pääsääntöisesti sataman infrastruktuuriin, toimintoihin ja niistä kerättyihin tapahtumatietoihin. Monesti kerätty data on sitä, mitä on mahdollista saada. Data on tallennettu vain keräysjärjestelmään eikä datan jatkokäyttöä ja hyödyntämistä ole välttämättä kovin tarkasti suunniteltu.

Ilman keskitettyä tietojen hallintaa data on todennäköisesti epätäydellistä ja sitä on hankalaa paikallistaa. Lisäksi voi olla vaikeaa tunnistaa, miten siihen on mahdollista päästä käsiksi ja onko data mahdollisesti myös helposti pirstoutuvaa, manipuloitavissa tai jopa tuhottavissa. Onko konteksti, johon data liittyy, selvästi kuvattu (metadata) ja ymmärrettävissä?

Sataman infrastruktuuridatan osalta päivitystaajuus ei ole keskeinen tekijä, mutta luotettavuuteen vaikuttaa datan säännöllinen ylläpito ja varsinkin nopea päivitys muutostilanteissa. Mikäli staattiseen infrastruktuuridataan yhdistetään dynaamista tapahtumadataa, koko informaation ajantasaisuusvaade nousee dynaamisen datan tasolle. Tällöin tunteja vanha data ei sovellu kaikkiin käyttötarkoituksiin. Pidemmät aikasarjat antavat tässäkin tapauksessa arvokasta tietoa esimerkiksi ennustamista ja etukäteissuunnittelua varten.

Esimerkiksi Digitraffic-järjestelmään kerättyä dataa voidaan perustellusti pitää luotettavana, koska se perustuu jatkuvaan valvontaan sekä olosuhteiden ja ympäristön mittaamisen pohjalta kerättävään dataan. Mahdollisia ongelmia voi syntyä, kun raakadataa käsitellään ja yhdistetään muuhun dataan ja siitä muodostetaan informaatiota. Tämä ei siis ole dataan itseensä liittyvä ongelma, vaan sen jalostamiseen ja käyttöön liittyvä haaste. Datan luotettavuutta arvioitaessa pitäisi siis huomioida kiinnittää sekä raakadatan sekä sen mahdollisen prosessoinnin jälkeiseen tilaan ja luotettavuuteen.

**Konteksti ja autenttisuus:** Tietojoukon sisältö ja jäljitettävyyys ovat keskeisiä tekijöitä sen alkuperäisyyttä ja luotettavuutta arvioitaessa. Kun datan ympärillä on asiayhteys, johon se liittyy, siitä muodostuu itsenäinen tietokokonaisuus. Kun data irrotetaan tietokokonaisuudesta, josta se on peräisin, asiayhteys menetetään. Datan avaaminen ilman siihen liittyvää kontekstin kuvausta (metadataa) saattaa vaarantaa datan arvon ja joskus tehdä siitä jopa käyttökeltovotonta. Kun avoimen datan julkaisussa kuvataan mm. tietolähteiden luonne

ja laatu, avoimesta datasta saadaan hyödyllistä raaka-ainetta esimerkiksi toiminnan tarkasteluun, avoimuuteen, vastuullisuuteen ja liiketoiminnan kehittämiseen.

**Datan säilytys:** sähköinen data saattaa ajan saatossa muuttua hankalakäyttöiseksi tai käyttökelvottomaksi. Uusien tallennusformaattien kehitys altistaa tietueet ja niiden yhtenäisyyden vanhentumiselle ja käyttökelvottomuudelle. Datavarantojen ylläpitoon ja suojaamiseen tulisikin niiden saatavuuden takaamiseksi kiinnittää huomiota.

On myös varmistuttava, ettei teknologian kehitys aiheuta sitä, että dataa ei joidenkin vuosien kuluttua pystytä käyttämään alun perin valitulta tietovälineeltä. Data on siis hyvä tallentaa ajan myötä uudemmille tietovälineille.

Hyvänä esimerkkinä ovat tietovälineet, kuten levykkeet, CD-levyt ja näitäkin vanhemmat magneettinauhut. Missään nykytietokoneessa ei ole levykeasemaa ja hyvin harvassa on CD-ROM-levyasema. Tämän hetken tallennuskapasiteettia ovat varsinkin pilvipalvelut. Myös tiedostoformaatit muuttuvat ja varminta onkin tallentaa tiedot jossakin kansainvälisessä standardimuodossa, jotka uusien formaattien kehittäjät ja ohjelmoijat yleensä ottavat paremmin huomioon.

**Digitaliset hakemistot:** luotettavat digitaliset hakemistot (Trusted Digital Repositories, TDR<sup>20</sup>) ovat kansainvälisesti hyväksytyt, teknologiariippumaton tapa varmistaa pitkäaikainen pääsy dataan ja tietueisiin sekä niiden käyttöarvon, luotettavuuden ja jäljitettävyyden turvaamiseen.

TDR:t suunnitellaan ja niitä hallitaan siten, että ne mahdollistavat pääsyn aitoon dataan ja tietueisiin sekä linkittävät aktiiviset ja ei-aktiiviset datasetit paperikopioon tai digitaalisiin tietokokonaisuuksiin. Ne tarjoavat kontekstin, keinot tulkita, liittää toisiinsa ja järjestää tietoja uudelleen sekä tukevat niiden siirtoa uusiin ohjelmisto- ja laiteympäristöihin.

## 8. Satamien kokemuksia datan avaamisesta

### 8.1. Pilottisatamien caset

Tämä osio täydentyy ja päivittyy DigiPort-hankkeen edetessä.

### 8.2. Muita esimerkkejä datan avaamisesta

**Suomen tie- ja katuverkon keskilinjageometria sekä tärkeimmät ominaisuustiedot** on koottu kansalliseen tietojärjestelmään Digiroadiin<sup>21</sup>. Digiroad tarjoaa kattavan, yhtenäisen, digitaalisessa muodossa olevan liikenneverkon kuvauksen. Aineisto mahdollistaa erilaisten reitinsuunnittelu-, navigointi-, matkailu- ja liikennelemaattisten palveluiden kehittämisen ja tuotteistamisen. Digiroad-tietopalvelusta vastaa Liikennevirasto. Kunnat ja ELY-keskukset vastaavat tietojen ylläpidosta yhteistyössä Liikenneviraston kanssa.

**Ajantasaista liikenne- ja olosuhdetietoa Suomen liikenneväyliltä** jaetaan liikenneviraston ylläpitämän Digitraffic-rajapintapalvelukokonaisuuden kautta. Digitraffic-palvelun tiedot kattavat kaikki kolme satamaliikenteen muotoa; tie-, rata- ja vesiliikenteen. Suurin osa tietolajeista saadaan Liikenneviraston ylläpitämistä tiedonkeruujärjestelmistä. Kaikki Digitraffic-palvelun kautta jaettava tieto on koneluettavaa avointa dataa ja on käytettävissä Creative Commons 4.0 käyttölisenssillä, mikä mahdollistaa uusien palveluiden ja ohjelmistojen kehittämisen.

---

<sup>20</sup> [ISO 16363 / TDR](#)

<sup>21</sup> <https://www.liikennevirasto.fi/avoindata/tietoaineistot#.W3LEfegzaUk>

**Sää ja sääolosuhteet** ovat tärkeitä tietoja meriliikenteelle. Ilmatieteen laitos tarjoaa avoimen lähdekoodin meteorologista tietoa "FMI Catalog" -verkkopalvelunsa kautta. FMI Catalog -palvelun tekninen toteutus noudattaa INSPIRE-direktiivin vaatimuksia. Toteutuksen sisältö on kuitenkin laajempi kuin sen vaatimukset. Palvelussa tarjottavat avoimet tiedot julkaistaan Open Geospatial Consortium -standardien mukaisina. Avoina datana julkaistaan sää-, meri- ja ilmastollisia havaintoja sekä reaaliaikaisia havaintoja, havaintojen ja ennusteiden aikasarjoja<sup>22</sup>. Tällaisia ovat esimerkiksi merijääennustedata, meriveden korkeustiedot ja aallonkorkeushavainnot suomea ympäröiviltä merialueilta.

Yleisölle suunnatut Ilmatieteenlaitoksen tavanomaiset sääpalvelut ja varoituspalvelut ovat aina maksuttomia. Hakupalvelun käyttö edellyttää kuitenkin rekisteröitymistä ja allekirjoittamalla rekisteröinnin käyttäjä hyväksyy myös lisenssisopimuksen<sup>23</sup>. Ilmatieteen laitos noudattaa asiakastiedostoja koskevaa lainsäädäntöä.

### 8.3. Hack the Port -kilpailu datan avaamisesta ja avoimen datan hyödyntämiskäytännöistä

Mussalon satamassa Kotkassa järjestettiin 9.-11.4.2019 Hack the Port -hackathon tapahtuma<sup>24</sup>. Hackathon on ongelmanratkaisutapahtuma, jossa opiskelijoista koostuvat tiimit pyrkivät kehittämään uusia, innovatiivisia ratkaisuja sataman haasteisiin. Hack the Portissa teemana oli satamatoimintojen sujuvoittaminen digitalisoidulla ja avoimen datan avulla.

Hack the Port -hackathonissa esiteltiin osallistujille viisi haastetta, joista neljä oli ennalta määriteltyjä ja yksi avoin, vapaavalintainen ratkaisuehdotus. Hack the Port -hackathonissa keskityttiin satamien toiminnan kehittämiseen.

Haaste 1: Sataman digitaalisen tilannekuvan kehittäminen

Haaste 2: Sataman maaliikenteen sujuvoittaminen digitalisoimalla

Haaste 3: Sataman turvallisuuden digitalisointi

Haaste 4: Sataman avointen tietolähteiden uhkien tunnistaminen ja ratkaisuehdotus

Haaste 5: Avoin haaste

Hackathon-haasteiden muotoilua varten järjestettiin HaminaKotkan ja Turun satamissa maaliskuussa 2018 työpajatilaisuudet. Tilaisuuksien nimi oli *"Mikä satamassa tökkii? Auttaisiko digi ja data?"* Niissä kerättiin osallistujilta tietoa heidän satamatoiminnoissa havaitsemiinsa kehityskohteisiin.

Osallistujia oli 20 ja he jakautuivat viiteen tiimiin. Opiskelijat edustivat eri koulutusohjelmia. Mukana oli niin tietoverkkojen, kyberturvallisuuden, rakennusalan ja palvelumuotoilun opiskelijoita kuin merikapteenikin. Monialaisuuden kautta on mahdollista yhdistää erilaista osaamista, jolloin voi syntyä "laatikon ulkopuolelta" -ratkaisuja.

#### 8.3.1. Hack the Port -kilpailun ratkaisuehdotuksia

Kilpailun voittajaratkaisussa ehdotettiin keinoja sataman liikenteen sujuvoittamiseksi ja aikasäästöjen saavuttamiseksi hyödyntämällä sähköpostilla ja mobiililaitteen välityksellä toimitettavaa satamaansaapumisilmoitusta ja -ohjetta rahdinkuljettajille.

Ratkaisussa satamaan lähetetään "passipyyntö", jossa kerrotaan ajoneuvon ja kuljettajan tiedot ja sähköpostiosoite. Paluupostissa kuljettaja saa sähköpostin, joka sisältää ajo-ohjeen karttoineen, tilauksen tiedot

<sup>22</sup> <http://en.ilmatieteenlaitos.fi/open-data-sets-available>

<sup>23</sup> <http://en.ilmatieteenlaitos.fi/open-data-licence>

<sup>24</sup> <https://www.xamk.fi/tapahtumat/hack-the-port/>



(tilauksen numero ja toimitusaika satamaan) sekä muuta tarvittavaa informaatiota. Tietoja voivat käyttää hyväkseen rahdinkuljettajien lisäksi satamassa toimivat organisaatiot.

Eräissä kilpailun ratkaisuehdotuksissa visioitiin myös sataman liikenteen turvallisuuden, sujuvuuden ja ympäristövaikutusten kehittämistä ja seuranta avoimen datan avulla. Ratkaisuehdotuksessa otettiin käyttötapaueksi tehtaalta satamaan saapuva ajoneuvoyhdistelmä, jonka tiedot tallennetaan järjestelmään kuljetuksen lähtöhetkellä ja välitetään tietoja hyödyntäville osapuolille etukäteen.

Sataman portilla ajoneuvo rekisteröidään mobiilisovelluksen avulla, jolloin ajoneuvo ja kuljettaja kirjautuvat sataman alueelle. Sataman alueella ajoneuvon käyttäytymistä (nopeus, reitti ja laskennalliset päästöt) voidaan seurata ja kirjata. Samoin ajoneuvon satamasta mahdollisesti ottama lasti voidaan liittää ajoneuvon tietoihin.

Satamasta lähtiessään ajoneuvo ja kuljettaja kirjautuvat vastaavasti järjestelmään, jolloin saadaan rekisteröityä ajoneuvon ja lastin (esim. kontit) lähtöaika satama-alueelta. Tämän perusteella voidaan mm. seurata satamassa käytettyä aikaa, jätettyä hiilijalanjälkeä, toimintojen sujuvuutta ja turvallisuutta.

Risteilymatkustajien pysäköintiongelmia Turun satamassa ratkottiin yhdessä kilpailuehdotuksista. Ongelmaa määriteltiin siten, että satamaan tullessaan risteilymatkustajien on hankala löytää pysäköintipaikkaa, koska he eivät useinkaan tunne sataman aluetta ja pysäköintipaikkojen sijaintia ennakolta. Lisäksi järjestelyt voivat olla muuttuneet ja opasteet eivät ole aina ajantasaisia.

Samassa kilpailuehdotuksessa ratkottiin myös HaminaKotkan sataman tilannetietojen ajantasaisuuden haasteita. Tilannetiedot ovat usein puhelimen ja sähköpostien varassa. Lisäksi ne eivät ole standardimuotoisia ja yhdenmukaisia, mikä voi aiheuttaa haittaa tietojen jakamiseen eri toimijoiden välillä. Tästä on seurauksena aikataulujen epäluotettavuutta ja taloudellisia menetyksiä.

Molempiin haasteisiin ehdotettiin ratkaisuksi konenäön avulla toteutettavaa tyhjen parkkipaikkojen tunnistamista ja tilannekuvan luomista liikenteestä satama-alueella. Nämä mahdollistaisivat aktiivisen opastuksen ja integraation auton navigaattoriin. Ratkaisuehdotus tuo mahdollisuuksia liikenteen valvontaan ja datan tuottamiseen sataman liikenteestä sekä mahdollistaisi saapuvan ajoneuvon ohjauksen perille asti ja pysäköintipaikan varaamisen ja maksamisen etukäteen. Järjestelmä mahdollistaisi myös parkkipaikkojen dynaamisen hinnoittelun. Kerättyä liikennedatata voitaisiin hyödyntää infraratkaisujen suunnittelussa ja dataa voidaan avata avoimen rajapinnan kautta hyödyntäjille, kuten matkanjärjestäjille viranomaisille ja satamassa toimiville yrityksille.