

Tavoitteena on saada kokemuksia datan avaamisesta normaalia tietojärjestelmien uusimissykliä nopeammin.

KUVA: A3sthetix / FLICR.com | CREATIVE COMMONS

## 2. Organisaation näkökulmia avoimuuteen

Datan avaamiseen ollaan Suomessa tarttumassa nyt laajalla rintamalla. Liikenne- ja viestintäministeriön on asettanut työryhmän edistämään julkisen informaation saatavuutta - valtiovarainministeriö kaksi (LVM & VM 2010). Valtakunnallinenkin media on tarttunut asiaan (HS 2.7.2010). Mitä tämä tarkoittaa yksittäisen organisaation kannalta?

Päätökset datavarantojen avoimesta ja maksuttomasta jakamisesta liittyvät taloudellisiin seikkoihin, vallankäyttöön, juridiikkaan ja tekniikkaan. Hyvin usein mukana on laaja toimijaverkosto, jolloin yhteisten päätösten tekeminen on aikaa vievää. Kansainvälisiä esimerkkejä - niin valtioiden, kaupunkien, kuin yksittäisten organisaatioidenkin - avoimeen dataan liittyvistä strategisista päätöksistä on runsaasti.

Oppaan kirjoittamisen aikaan yleinen ilmapiiri hallinnon datan avaamista kohtaan on nopeasti muuttunut positiivisemmaksi, mutta aihetta ei ole suoraan mainittu vielä monienkaan suomalaisten organisaatioiden strategioissa tai muissa julkilausutuissa tavoitteissa. Yleisemmällä tasolla esimerkiksi kansallisessa tietoyhteiskuntastregiassa (VNK 2006) tai EU-lainsäädännön kautta (Inspire ja PSI) aihe on kuitenkin jo heijastunut useimpien julkishallinnon organisaatioiden toimintaan.

Avoimen datan julkaiseminen tukee monien julkishallinnon organisaatioiden strategisia tavoitteita, kuten edistää kansalaisten osallistumismahdollisuuksia ja yritysten liiketoimintamahdollisuuksia. Useimmiten datan avoin ja maksuton tarjoaminen linkittyy suoraviivaisimmin viestintä- ja tietohallintostrategioihin. Tämä kuitenkin riippuu hyvin paljon organisaatiosta. Esimerkiksi Tilastokeskuksen tehtävä on tiedon tuottaminen, jolloin avoimen datan tarjoaminen liittyy suoraan sen strategiaan. Sen sijaan esimerkiksi TEKESin ja Finpron kaltaisten organisaatioiden tavoitteet liittyvät suomalaisen yritystoiminnan edistämiseen. Tällöin suhde dataan on välillisempi.

Yleensä pyrkimyksenä on, että organisaation muutoinkin tuottamille tietovarannoille saadaan avoimuuden myötä lisää käyttöä, joka tavalla tai toisella hyödyttää myös organisaatiota. Avoimen datan ekosysteemi ei kuitenkaan toimi tarkan vastavuoroisuuden perusteella, vaan hyödyt tulevat osittain epäsuorasti ekosysteemin kokonaisuuden kautta. Esimerkkejä epäsuorista hyödyistä ovat organisaation tunnettavuuden kasvu ja sen tarjoamien palveluiden kysynnän kasvaminen.

Yksittäiset julkishallinnon organisaatiot ovat datan avaamisessa aloitteen tekijän asemassa. Useimmissa organisaatioissa resurssit datan avaamiselle

eivät kuitenkaan vielä ole merkittäviä. Tässä luvussa hahmottelemme organisaation näkökulmasta osallistumista avoimen datan ekosysteemiin. Liikkeelle kannattaa lähteä kevyesti, mutta tavoitellen näkyviä muutoksia ja todellista käyttöä datalle. Kun omat tietovarannot on identifioitu, voidaan niistä poimia helpoimmat avattavaksi ensin. Näiden alhaalla roikkuvien hedelmien kanssa päästään keräämään kokemuksia datan avaamisesta. Ensimmäisten pilottien perusteella voidaan organisaatiolle muodostaa tietovarantojen avoimuutta edistävä strategia, jonka mukaisesti rekistereitä ja rajapintoja järjestelmällisesti avataan.

Tässä luvussa esiteltävät datan avoimuudet mittarit (2.2) ja viitteellinen datavarantojen inventaario- ja julkaisuprosessi (2.3) muodostavat rungon seuraaville luvuille, joissa syvennetään maksuttomuuden, laillisen ja teknisen uudelleenkäytettävyyden, sekä datan löydettävyyden teemoja. Datan avaamishankkeessa on hyvä olla mukana laaja edustus organisaation osaamista, mukana tulisi olla ainakin informaatio-, laki-, talous-, tietotekniikka-, viestintä ja hallinto-osaamista sekä lisäksi päätösvaltaisia ihmisiä.

## 2.1 Datavarantoja toimijoita ja rooleja

Potentiaalisia toimijoita avoimen datan ekosysteemissä on niin julkishallinnon organisaatioissa, yrityksissä, järjestöissä kuin kansalaisten parissa. Julkishallinnolla on merkittävä rooli datan tuotannossa ja siksi olisi oleellista, että mahdollisimman suuri osa julkishallinnosta on mukana. Julkishallinnon datasta puhuttaessa tulee yleensä ensimmäisenä mieleen tunnetut perusrekisterit, kuten väestörekisteri, kiinteistörekisteri jne. sekä kenties tilastotiedot ja sää tiedot. Myös kartta-aineistot ja paikkatieto on melko laajasti tunnettua ja jotkut muistavat mainita erilaiset liikennetiedot. Hallinnon data-aineistoja ei Suomessa ole kattavasti kartoitettu, mutta niiden määrä on eri arvioiden mukaan kuitenkin hämmästyttävän suuri.

Case: Dataa on kaikkialla – puurekisteri ja kirjaston tietokannat Käytännön sovellusesimerkki New Yorkista, ja ajatuskulku kotimaisiin kirjastoihin liittyen osoittavat, että julkista dataa on kaikkialla kunhan sen olemassaolon huomaa.

- 1) Brett Camperin toteuttama Puut lähelläsi (Trees Near You) -palvelu sai kunniainiminnan New Yorkin kaupungin avoimen datan hyödyntämiskilpailussa 2010. Ilmainen iPhone-sovellus kertoo käyttäjälle sijainnin perusteella tietoja yli 500 000 elävästä puusta New Yorkin kaupungissa. Sovellus yhdistelee puhelimesta saatavaa GPS-sijaintitietoa, Wikipedia-artikkeleita puulajeista ja kaupungin avoimesti julkaisemaa puurekisteriä. Tämä on hyvä esimerkki siitä, kuinka paljon potentiaalisesti hyödyllistä dataa on olemassa mutta jonka olemassaoloa ei yleensä tule ajatelleeksikaan.

- 2) Kirjastojen tietokannoista voisi mainita aineistojen hyvät bibliografiset tiedot: tekijä, teoksen nimi, teoksen kansikuva, sisällysluettelo ja sisällönkuvailu. Useilla maakuntakirjastoilla on myös aluetietokantoja, esim. Tampereen maakuntakirjasto ylläpitää Pirkitta-aluetietokantaa, johon on tallennettu muun muassa kotimaisten aikakauslehtien artikkeliviitteitä vuodesta 1983 lähtien ja digitoituja näköispainoksia pienpainatteista. Tällaiselle paikalliselle datalle voisi löytyä soveltamisalueita, jos tietokanta sen sallisi.

Luonnollisesti jokaisesta lainatusta kirjasta jää myös jälki tietokantaan. Nyt tietokannoissa on mukana lainaajia koskevaa tietosuojan alaista dataa, mutta ilman hnekilötietoja (anonymisoidut) olevat lainaustilastot, kuten lainatuimmat aineistot jne., olisivat mielenkiintoisia. Asiakkaalle itselleen olisi mielenkiintoista saada oma lainaushistoriansa käyttöön, hän voisi linkittää sen vaikka verkossa toimivaan lukupiirisovellukseen, jolloin hän pääsisi vaihtamaan kommentteja muiden ihmisten kanssa, jotka ovat lukeneet samoja kirjoja.

Julkishallinnon datavarantoja voidaan jäsentää esimerkiksi tuottajaorganisaatioiden, aineistojen sisällön tai datan oletettujen käyttötarkoitusten perusteella. Aiemmissa keskustelunavauksissa tarkastelua on kentän laajuuden takia yleensä rajattu esimerkiksi vain joihinkin seuraavista: paikkatieto (MMM 2005), organisatorisesti, esim. valtion budjettitalouden piiriin kuuluvien organisaatioiden tuottama data (Kuronen 1998). Tässä oppaassa laajennamme tarkastelua koskemaan periaatteessa kaikkea lain mukaisesti julkista, hallinnon tuottamaa tietoa.

On olennaista, että organisaatioissa tunnistetaan omat tietovarannot ja niistä jaetaan tietoa ja viestitään. Valtion tasolla on meneillään tietoarkkitehtuurin määrittelytyö (VM 2009), jonka puitteissa on tehty alustava kaavio suomalaisen datavarantojen kokonaiskenttää: *Tietomaaston kartoitusta – merkittäviä tietokokonaisuuksia* (Liite 3).

Vaikka oppaassa keskitytään nimenomaan julkishallintoon merkittävänä informaation tuottajana, niin ekosysteemin näkökulmasta ei ole olennaista, kuka tuottaa ja kuka käyttää informaatiota. Nykyään tilanne saattaakin olla jo niin, että yksityiset yritykset tuottavat lähes yhtä paljon dataa kuin julkishallinnon organisaatiot. Enenevässä määrin myös yksittäiset kansalaiset osallistuvat datan tuotantoon ja sen jalostamiseen informaatioksi ja tiedoksi. Avoimen datan ekosysteemissä hallinnon organisaatiot, kansalaiset ja yritykset ovat kaikki paitsi tiedon käyttäjiä myös tiedon tuottajia – usein molempia samaan aikaan. Ekosysteemi-ajattelussa toimijoita ovat paitsi organisaatiot, niin myös yksittäiset ihmiset, jotka ovat keskenään vuorovaikutuksessa ja muokkaavat ekosysteemin käytäntöjä.

Moniin hallinnon prosesseihin, kuten esimerkiksi lakien ja asetusten valmisteluun, liittyy paljon organisaatorajat ylittävää viestintää ja tiedonvaihtoa.

Lainvalmistelu saattaa esimerkiksi pitää sisällään Tilastokeskuksen tietojen hyödyntämistä ja budjettivaikutusten arvioimista valtionvarainministeriössä. Käytännöllisenä lähtökohtana voidaan pitää, että jokainen julkishallinnon organisaatio todennäköisesti sekä a) tuottaa uutta dataa että b) jalostaa, c) käsittelee ja d) hyödyntää jonkun muun tuottamaa dataa. Vaikka toimijakenttä on moninainen, eikä ole selkeästi osoitettavissa julkisen datan tuottajaorganisaatioita, niin toimijoiden roolit suhteessa dataan on kuvattavissa tarkasti (taulukko 2.1).

*Taulukko 2.1. Esimerkkejä toimijoiden rooleista suhteessa dataan.*

Tallentaja raakadatan kerääminen ja tallentaminen Jalostaja raakadatan käsittely ja muokkaaminen Aggregaattori datan yhdistely ja koostaminen eri lähteistä Harmonisoija eri lähteistä tulevien tietojen yhdenmukaistaminen ja yhteismitallistaminen (samalta näyttävä asia myös tarkoittaa samaa) Päivittäjä tietojen päivittäminen Julkaisija datan julkaiseminen Rekisterinpitäjä datavarannon ylläpitovastuu Sovelluskehittäjä datan loppukäyttäjänä datan hyödyntäminen osana palvelua Tulkitsija datan loppukäyttäjänä datan tulkitseminen, esim. tutkija, yritys tai demokratiaaktivisti Dataan pohjautuvien palveluiden (ks. luku 2) käyttäjät ihmiset yritykset ja organisaatiot, jotka käyttävät avoimen datan päälle tehtyjä sovelluksia ja tulkintoja Myöhemmin tässä luvussa esitettävän organisaation tietovarantojen inventaarion yhteydessä identifoidaan aineistot, joita organisaatiolla on hallussaan. Yllämainittu rooli-jaottelu sekä liitteen 3. julkishallinnon tietomaaston kartoitus on tarkoitettu avuksi inventaarioon.

## 2.2 Datan avoimuuden mittareita

Keskustelunavausten, raporttien ja hankkeiden myötä avoimuudesta on nopeasti tullut keskeinen sana hallinnon sanastoon myös kun puhutaan tietojärjestelmistä ja tietovarannoista. Avoimuuskeskustelussa puhutaan samaan aikaan avoimesta lisensoinnista, teknisistä rajapinnoista, formaateista, metadatasta, tiedon harmonisoinnista, hallinnon läpinäkyvyydestä, ihmisten osallistumismahdollisuuksista, julkisuusperiaatteesta, uudelleenkäytettävyydestä ja koneluettavuudesta. Ei ole aivan poikkeuksellista, että eri osapuolet tarkoittavat eri asioita. Aiheen laajuuden ja nopean kehityksen vuoksi väärinymmärrykset ja liika yleistäminen ovat yleisiä. Mitä oikeastaan tarkoitetaan avoimella julkishallinnon datalla?

Lainsäädännön mukaan julkisella datalla tarkoitetaan sellaisia aineistoja, jotka eivät kuulu yksityisyydensuojan piiriin – kuten henkilörekisterit – tai ole muusta syystä laissa arkaluontoisiksi tulkittavia, kuten kansalliseen turvallisuuteen liittyvät aineistot.

Aineiston avoimuus ei ole mustavalkoinen joko-tai-kysymys, vaan siihen vaikuttaa ainakin aineiston tekninen saavutettavuus, löydettävyys, maksullisuus ja lisensointiehdot. Kaikkien aineistojen kohdalla äärimmäinen avoimuus ei

ole edes tavoiteltavaa. Monesti aineiston uudelleenkäytettävyyttä voidaan merkittävästi parantaa ilman haittoja tekemällä ratkaisuja, jotka tavalla tai toisella lisäävät avoimuutta (esimerkiksi muuttamalla lisensointietoja sallivammiksi tai tarjoamalla aineistot täysin veloitusetta).

Data-aineiston avoimuutta voidaan arvioida vertaamalla sitä avoimen datan kriteeristöön. Seuraava määritelmä pohjautuu Open Knowledge Defition -määritelmään (OKD). Julkisen datan täysin avoimella löydettävyydellä, saatavuudella ja uudelleenkäytettävyydellä tarkoitetaan seuraavaa:

*Julkisella datalla tarkoitetaan aineistoja, joita eivät koske yksityisyydensuoja-, turvallisuus- tai muut rajoitteet, jotka tekisivät datasta salassa pidettävää tai arkaluontoista.*

- *Aineisto on löydettävissä ja kokonaisena saatavilla siten, että se voidaan ottaa laillisesti käyttöön välittömästi ilman viranomaisen avustusta esimerkiksi kello kolme yöllä.*
- *Aineisto on vapaasti tarjolla internetissä. Kuka tahansa saa lukea, ladata sen omalle koneelleen, kopioida, muokata, jakaa, tulostaa, etsiä, linkittää dataan, indeksoida, antaa data tietokoneohjelman syötteeksi tai käyttää sitä mihin tahansa muuhun lainmukaiseen toimintaan ilman taloudellisia, juridisia, teknisiä tai käytännöllisiä rajoitteita.*
- *Käyttöehdoilla turvataan aineiston käyttäjälle varmuus aineiston alkuperästä ja aineiston tuottajalle oikeus tulla asianmukaisesti nimetyksi sen hyödyntämisen yhteydessä. Muita aineiston uudelleenkäyttöä ja jakelua rajoittavia ehtoja ei ole ja aineisto on saatavilla ilman rekisteröitymisvaatimuksia.*
- *Aineiston yhteyteen (esim. metatietoihin) sisällytetään tiedot aineiston alkuperästä ja käyttöehdoissa kerrotaan selkeästi aineiston vastaanottajalle, että hänellä on laillinen oikeus ja häntä jopa kannustetaan uudelleenkäyttämään aineistoa parhaaksi katsomallaan tavalla.*

Koska täysin avoimen datan määritelmä harvoin kaikilta osin täyttyy tai on edes tarpeen täyttyä, on määritelmä purettu osiin: löydettävyys (2.2.1), kokonaisuus (2.2.2), käyttöehtojen tasa-arvoisuus (2.2.3), alkuperäisyys ja ajantasaisuus (2.2.4), laillinen ja vapaa uudelleenkäytettävyys (2.2.5), maksuttomuus (2.2.6), koneluettavuus (2.2.7), formaatin avoimuus (2.2.8) ja ymmärrettävyys (2.2.9). Tämä osiin purettu avoimen datan määritelmä toimii akselistona, jonka avulla voidaan arvioida yksittäisen julkisen datalähteen avoimuutta ja tunnistaa missä asioissa sitä kenties voitaisiin helposti lisätä.

### 2.2.1 Löydettävyys

**Avoim:** Aineiston olemassaolo ja sijainti on yleisesti tunnettu. Aineisto ja sen uudelleenkäytön sallivat lisensointiehdot ovat helposti löydettävissä internetistä

sekä ihmisille että hakuroboteille.

**Suljettu:** Aineisto on vain viraston operatiivisessa järjestelmässä, eikä sen olemassaolosta ole ulkopuolisilla lainkaan tietoa.

Google Maps -rajapinta tai Wikipedian sisältö ovat esimerkkejä aineistoista, joiden olemassaolo ja laillinen käytettävyys on yleisesti tiedossa. Aineiston löydettävyyttä voidaan parantaa lisäämällä se hyvin ylläpidettyyn datakatalogiin, optimoimalla datavarantojen kuvailutiedot hakukoneita varten ja julkaisemalla data linkitetyn datan paradigman mukaisesti. Yleistä tunnettuutta voidaan vahvistaa tiedottamalla aktiivisesti potentiaalisille uudelleenkäyttäjille datavaranon olemassaolosta verkossa, julkaisuissa ja tapahtumien yhteydessä. (ks. luku 6: Avoimen datan infrastruktuuri)

### 2.2.2 Kokonaisuus

**Avoim:** Data on kokonaisuudessaan vapaasti ladattavissa internetistä. Saavutettavuutta ja datan käyttömahdollisuuksia ei rajoiteta epäsuorasti, tarjoamalla pääsy vain osaan tietokannasta kerrallaan.

**Suljettu:** Aineistosta on saatavilla avoimesti vain osa ja koko aineiston saaminen edellyttää erillistä sopimusta.

Tyypillinen tapaus, jossa aineiston kokonaisuutta rajoitetaan tarkoituksella tai tahattomasti on sellainen, jossa aineisto tarjotaan vain kyselyrajapinnan kautta, eikä aineistoa ole mahdollista ladata kokonaan. Mikäli aineisto on saatavilla kokonaisuudessaan tarkoittaa se, että kenen tahansa on teknisesti mahdollista myös ryhtyä jakamaan dataa itseään ja muita varten. Kokonaisuuden rajoittaminen voi olla keino varmistaa aineiston loukkaamattomuus (integriteetti) ja estää rinnakkaisten kopioiden syntyminen aineistosta. Toisaalta kokonaisuuden rajoittaminen estää datan kattavaan analyysiin perustuvan käytön ja aiheuttaa kyselyrajapinnan kuormitusta, joka voitaisiin välttää tarjoamalla oma kopio. (ks. luku 5: Tekninen valmistautuminen)

### 2.2.3 Käyttöehtojen tasa-arvoisuus

**Avoim:** Aineisto on saatavilla kaikille ja kaikkiin laillisiin käyttötarkoituksiin yhtä helposti. Datan luovutuksessa ei aseteta käyttäjiä eikä käyttötarkoituksia eriarvoiseen asemaan, vaan muut hallinnon toimijat, kansalaiset, yritykset ja myös ulkomaalaiset toimijat ovat tasa-arvoisia.

**Suljettu:** Aineiston tarjoamista on rajoitettu käyttäjän tai käyttötarkoituksen mukaan, esimerkiksi vain tutkimus ja tuotekehityskäyttöön tai ainoastaan epäkaupalliseen käyttöön tai ainoastaan suomalaisille.

Käytännössä tasa-arvoisuus toteutuu, mikäli aineisto on saatavissa verkosta ilman rekisteröintivaatimuksia, jolloin kuka tahansa voi ottaa sen käyttöönsä

standardinmukaisilla lisenssiehdoilla. Lisenssi ei estä ketään käyttämästä dataa jollakin määrättyllä käyttöalueella. Erityisesti myös kaupallinen käyttö on sallittua, sillä kaupallisten toimijoiden toivotaan liittyvän avoimen datan ekosysteemiin. Tasa-arvoisuus tarkoittaa ennakoidusta kontrollista luopumista. Dataa saa käyttää myös taitamattomasti ja poliittisiin tarkoituksiin. (ks. luku 3: Lupa julkaista ja käyttää)

#### 2.2.4 Alkuperäisyys ja ajantasaisuus

**Avoin:** Aineisto on tarjolla alkuperäisessä muodossaan ja alkuperäisellä tarkkuusasteella silloin, kun se ei uhkaa yksilöiden tietosuojaa. Aineisto julkaistaan lähimpänä sen syntypaikkaa (alkuperäisen tuottajan toimesta) ajantasaisesti suhteutettuna aineiston sisällön muuttumisnopeuteen.

**Suljettu:** Aineisto julkaistaan viivästetysti, alkuperäistä pienemmällä tarkkuudella tai ainoastaan yhdistetyssä muodossa muiden aineistojen kanssa.

Raakadatan lisäksi myös yhdistettyjä ja jalostettuja muotoja datasta voidaan asettaa vapaasti saataville. Joissain tapauksissa yleistämällä ja laskemalla tarkkuustasoa voidaan muutoin yksityisyydensuojan kannalta arkaluontoinen aineisto saattaa julkiseksi. Anonymisointi ja yleistäminen täytyy tehdä kuitenkin erityisellä huolella. (ks. luku 5: Tekninen valmistautuminen)

#### 2.2.5 Laillinen ja vapaa uudelleenkäytettävyys

**Avoin:** On olemassa standardikäyttöehdot, joiden puitteissa aineisto on vapaasti uudelleenkäytettävissä kaikille riippumatta käyttäjästä tai käyttötarkoituksesta. Mahdollisia yksityisyydensuojan, turvallisuuteen liittyviä rajoituksia lukuunottamatta datan käytölle ei ole asetettu muita juridisia rajoituksia. Lisenssintietoa on esitetty selvästi ja läpinäkyvästi siten, että ne kannustavat uudelleen käyttöön. Aineistoon mahdollisesti kohdistuvista tekijänoikeuksista on luovuttu (Waiver of Rights) ja se kerrottu suoraan aineiston uudelleenkäyttöön kannustavissa käyttöehdoissa.

**Suljettu:** Aineisto on lisensoitu uudelleenkäyttöä rajoittavalla lisenssillä, siihen kohdistuu tekijänoikeuksia, tai sen uudelleenkäytön sallimista ei ole suoraan esitetty missään.

Salliva lisenssintietoa voidaan toteuttaa esimerkiksi Creative Commons, Open Database -lisenssillä tai vastaavalla. Aineistoon mahdollisesti kohdistuvista tekijänoikeuksista tulisi avoimen julkisen datan osalta luopua, esimerkiksi Creative Commons nolla -lisenssillä, jolloin vältetään epäselvyyksiltä koko jatkojalostusketjussa. Yleisin haastatteluissa esiin noussut toive käyttöehtoihin liittyen on halu tietää ketkä aineistoa käyttävät ja mihin. Hyvin usein ei ollut mitään varsinaisia tarpeita rajoittaa käyttötarkoituksia, mutta ne haluttiin kuitenkin tietää oman toiminnan kehittämisen näkökulmasta. Aineistojen käytön seuranta voi-

daan tehdä myös ilman allekirjoitettavia sopimuksia ja rajoittavia käyttöehtoja, esimerkiksi verkossa tapahtuvan käyttjärekisteröinnin avulla. (ks. luku 3: Lupa julkaista ja käyttää)

### 2.2.6 Maksuttomuus

**Avoim:** Data on saatavissa täysin maksutta.

**Suljettu:** Dataa tarjotaan maksuperustaisesti ja myynnistä saaduilla tuotoilla katetaan dataa tuottavan organisaation muita toimintakuluja.

Enintään irtiottokustannuksilla tarjottavaa aineistoa voidaan pitää melko avoimena. Hyvin usein kuitenkin aineiston ylläpidon ja tuotannon kustannukset ovat moninkertaiset suhteutettuna irtiottokuluina perittäviin maksuihin. Pienikin maksu rajoittaa aineiston käyttöä merkittävästi siihen liittyvän vaivaloisuuden ja sopimusten takia. Voi olla, että suurin osa irtiottokustannuksista muodostuu itseasiassa sen laskuttamiseen liittyvästä byrokratiasta, jolloin irtiottokustannuksien perimistä ei voida pitää perusteltuna. Mikäli erityisestä syystä peritään maksuja, tulisi maksu olla mahdollista suorittaa internetissä ja saada aineisto käyttöön välittömästi ilman viranomaisen työtä. (ks. luku 5: Avoimen datan taloudesta)

### 2.2.7 Koneluettavuus

**Avoim:** Datavarannolla on pysyvä sijainti internetissä ja sitä pääsee lukemaan automatisoidusti ja ohjelmallisesti. Data on riittävässä määrin strukturoitu automatisoidun prosessoinnin mahdollistamiseksi. Myöskin lisensointiehdot ovat koneluettavia, ne voidaan hyväksyä verkossa ja aineisto saadaan käyttöön ilman viranomaisen työtä.

**Suljettu:** Data julkaistaan täysin epärakenteellisessa muodossa siten, että sen tulkinta on mahdollista vain ihmislukijalle, esim. pdf-dokumentit tai html-sivut.

Peukalosääntönä koneluettavuudelle voidaan pitää sitä, että etevä ohjelmoija kykenee verrattain lyhyessä ajassa tekemään ohjelman, joka automaattisesti hakee datan internetistä, lukee sen ohjelman muistiin ja tulostaa uudelleen muokattuna vaikkapa iPhoneen näytölle. Mikäli aineisto ei missään vaiheessa ole koneluettavassa muodossa, on sen saattaminen sellaiseksi varsin työlästä. Usein kuitenkin esimerkiksi organisaation verkkosivuilla tarjotaan aineistoja epäkoneluettavasti HTML-muodossa, mutta samat aineistot ovat jossain toisessa järjestelmässä myös koneluettavassa muodossa, jolloin niiden julkaiseminen on jo paljon helpompaa. (ks. luku 5: Tekninen valmistautuminen)



### 2.2.8 Formaatin avoimuus

**Avoim:** Aineisto on saatavilla käyttökelpoisessa ja muokattavassa formaatissa, jonka määrittely ja kehitys eivät ole minkään yksittäisen yrityksen tai vastaavan hallussa.

**Suljettu:** Data on tarjolla ainoastaan formaatissa, joka on jonkin yrityksen hallinnoima ja jonka hyödyntämiseksi on pakko käyttää saman yrityksen ohjelmistoja.

Tämä voidaan saavuttaa tarjoamalla aineisto avoimessa formaatissa, esimerkiksi sellaisessa, jonka spesifikaatio on julkisesti ja vapaasti saatavilla ja joka ei aseta rahallisia tai muita rajoitteita formaatin käytölle. Samaa dataa kannattaa mahdollisuuksien mukaan tarjota useammissakin formaateissa. Aina avoimen formaatin käyttöön ei ole realistisia mahdollisuuksia. Esimerkiksi jotkut paikatietojärjestelmät käyttävät valmistajakohtaisia formaatteja, jolloin avoimiin formaatteihin voidaan siirtyä vain järjestelmäuudestuksen yhteydessä. (ks. luku 5: Tekninen valmistautuminen)

### 2.2.9 Ymmärrettävyys

**Avoim:** Data on kuvailtu ja dokumentoitu kattavasti ja selkeästi.

**Suljettu:** Aineisto on tarjottu formaatissa, jonka käyttöä tai sisältöä ei ole kuvailtu uudelleenkäytön edellyttämällä tavalla.

Aineiston ymmärrettävyyttä ja sitä kautta uudelleenkäytettävyyttä voidaan merkittävästi parantaa metadatan, dokumentaation, käyttöesimerkkien ja laatumääritelmien avulla. Hyvän dokumentaation ainoa haittapuoli on sen tekemisen työläys. Ohjeistaminen saattaa muodostua merkittäväksi aineistojen julkaisemisen hidasteeksi. Toisaalta dokumentaatio voi joskus olla hyvinkin kevyt ja sitä voi parantaa myöhemmin. Esimerkiksi sarkainerotellussa tiedostossa riittää, että sarakeotsikot ovat tiedostossa mukana. (ks. luku 5: Tekninen valmistautuminen)

## 2.3 Datavarantojen inventaario- ja julkaisuprosessi

Hyvin kuvailtu laadukas data luotettavasti toimivien rajapintojen kautta jaeltuna johtaa avoimen datan suurimpiin hyötyihin. Datan avaaminen vaatii kuitenkin aikaa ja resursseja, eikä kaikkea voida rakentaa kerralla valmiiksi. Datan jakamisessa kannattaa etsiä pieniä toimia, joilla datan avoimuutta voidaan merkittävästi lisätä.

Esitämme tässä datan avaamiseen vaiheittaisen ja käytännön kokeilujen kautta ohjautuvan etenemismallin. Tavoitteena on, että datan avoimesta jakelusta ja ekosysteemiin osallistumisesta saadaan kokemuksia aiemmin, kuin mitä tietojärjestelmien uusimissykli toisi mukanaan. Nähdään mille datalle on eniten

kysyntää ja saadaan palautetta siitä, missä formaateissa ja millaisen rajapinnan kautta mitäkin kannattaa jakaa. Prosessi on tarkoitettu viitteelliseksi ja ennen kaikkea herättämään ajatuksia.

Datan avaamisessa kannattaa välttää turhia ennakko-oletuksia datan käyttötarkoituksesta ja potentiaalisista hyödyntäjistä. Joillekin aineistoille voi olla piilevää kysyntää, joka paljastuu vasta kun aineisto saa julkisuutta. Esimerkiksi paikkatietolainsäädännöstä käyty keskustelu on lisännyt paikkatietoaineistojen tunnettuutta ja sitä kautta niiden kysyntää. Priorisointimielessä on loogista selvittää mille tietovarannoille on eniten kysyntää. Mielikuvien sijaan on kuitenkin parempi nojautua kokeiluihin, kannustaa datan hyödyntämiseen ja seurata mitä tapahtuu. Yksikin julkisuutta saanut sovellus saattaa johtaa aineiston kysynnän lisääntymiseen ja muuttaa tilannetta lyhyessä ajassa merkittävästi.

Teknisesti tietovarantojen julkaisu liittyy organisaation tietojärjestelmien ja etenkin tietoarkkitehtuurin kehittämiseen. Aineistojen julkaisua voidaankin ajatella tietoarkkitehtuurin kehittämisen tukitoimintona eikä niinkään arkkitehtuuriprojektin lopputuloksena. Sen sijaan, että luvataan avoimet rajapinnat seuraavan järjestelmäuudistuksen yhteydessä, lähdetäänkin siitä, että ensin julkaistaan organisaation tietovarantojen inventaarion tulokset. Sen jälkeen julkaistaan joitain kiinnostavia tietoaineistoja ladattavaksi sellaisena, kuin ne ovat. Datan julkaisusta saadut kokemukset voivat toimia tärkeänä syötteenä niin aineistojen harmonisointiin kuin yleisen tietojärjestelmien kehitystyöhönkin.

Vaiheittaisen prosessin (kuva 2.1) jokaisella kierroksella syvennetään organisaation omien tietoresurssien tuntemusta, julkaistaan uutta materiaalia ekosysteemin käytettäväksi ja kerrytetään organisaation kokemuksia avoimeen dataan liittyvästä toiminnasta. Prosessin osia ovat: a) aineistojen analysointi, b) aineiston julkaiseminen ja c) datan avaamisesta oppiminen.

**Ensimmäinen kierros: Datasta ilmoittaminen** Datan avaamisesta ja organisaation hallussa olevista aineistoista ilmoittaminen pyritään tekemään mahdollisimman nopeasti ja kevyesti. Tässä vaiheessa vastataan kysymyksiin:

\* 1) Mitä aineistoja meillä ylipäättään on? \* 2) Mitkä niistä ovat julkisia? \* 3) Kuinka avoimia eri aineistot nykyisin ovat? (ks. mittareita datan avoimuudesta)

**Toinen kierros: Pilottiaineistojen julkaiseminen** Kun omat tietovarannot on tunnistettu, niistä voidaan poimia helpoimmat avattaviksi ensin. Seuraavalla iteraatiokierroksella pilotoidaan teknisesti, juridisesti ja taloudellisesti helpoimmin avattavien aineistojen julkaisua. Tässä vaiheessa vastataan kysymyksiin:

\* 4) Mitkä aineistot ovat teknisesti helposti avattavissa? \* 5) Minkälaisin käyttöehdoin niitä halutaan julkaista? \* 6) Ketkä organisaatiossa vastaavat eri aineistojen sisällöstä ja ketkä teknisistä järjestelmistä?

**Kolmas kierros: Dokumentointi ja käyttötapaukset** Julkaistujen aineistojen määrä voidaan kierros kierrokselta lisätä, mutta samalla on hyvä lisätä

ja selkiyttää aineistojen dokumentaatiota uudelleenkäyttäjää varten sekä koota aineistojen käyttötapauksia ja muuta tietoa organisaation sisältä. Tässä vaiheessa vastataan kysymyksiin:

\* 7) Miten tieto on kuvailtu sisällöllisesti ja dokumentoitu teknisesti? \* 8) Millaisia tiedontarpeita, käyttäjäryhmiä ja käyttötapauksia eri aineistojen nykyiseen käyttöön liittyy? \* 9) Minkälaisia tiedon saatavuuteen ja käytettävyyteen liittyviä pyyntöjä ja tarpeita on noussut esiin organisaation sisältä ja ulkopuolelta?

**Neljäs kierros: Tietoarkkitehtuuri ja sanastotyö** Ensimmäisten pilottien kokemuksella voidaan organisaatiolle muodostaa tietovarantojen avoimuutta edistävä strategia, jonka mukaisesti rekistereitä ja rajapintoja järjestelmällisesti avataan. Sen jälkeen on mahdollista kehittää tietovarantojen käyttöä myös yli organisaatorajojen. Tässä vaiheessa vastataan kysymyksiin:

\* 10) Minkälaisia organisaatorajat ylittäviä tiedon harmonisointi- ja yhdenmukaistamistarpeita ja mahdollisuuksia on? \* 11) Minkälaisiin järjestelmäuudistuksiin tietovarantojen avoimuuden parantaminen liittyy? \* 12) Miten organisaation käyttämiä datavarantoja voitaisiin paremmin organisoida yhdessä muiden toimijoiden kanssa?

Kuten jo aikaisemmin mainittiin, niin prosessi on viitteellinen ja toteutus saat-  
taa vaihdella organisaatiottain. Lähtöpiste on kuitenkin selvä. Datan avaami-  
nen lähtee aina omien datavarantojen tunnistamisesta ja nykytilan arvioimis-  
esta. Tämän jälkeen on mahdollista osallistua avoimen datan ekosysteemiin  
– ensin projektimielessä kokeillen ja sen jälkeen sisällyttäen avoin data osaksi  
organisaation pysyvää toimintatapaa. Liikkeelle voidaan lähteä pienestä ja ker-  
tyneet kokemukset antavat aineksia organisaation omaan avoimen datan strate-  
giaan.

## A) Aineistojen analysointi

- Ensimmäinen kierros: lista organisaation hallussa olevista aineistoista
- Toinen kierros: analyysi aineistoista (tekniikka, juridiikka, vastuut)
- Kolmas kierros: haastattelut ja käyttötapaukset tietoresurssien käytöstä
- Neljäs kierros: tietoarkkitehtuurin kehittämissuunnitelma

## B) Aineistojen julkaiseminen

- Ensimmäinen kierros: julkaistaan organisaation tietoa-aineistojen rekisteri (Information Asset Registry IAR)
- Toinen kierros: julkaistaan pilottiaineistot raakadatamuodossa (ns. ”dumppijulkaisu”)

- Kolmas kierros: määritellään rajapinta ja julkaistaan dokumentaatiota
- Neljäs kierros: tehdään hakupalvelu tai dataportaali

### C) Datan avaamisesta oppiminen

- Ensimmäinen kierros: seurataan tilastoja aineistojen käytöstä ja latauksista
- Toinen kierros: ensimmäiset sovellukset, sana leviää
- Kolmas kierros: tavataan käyttäjiä esimerkiksi aineiston käyttäjien työpajassa
- Neljäs kierros: organisaation datavarantoja käytetään ja jalostetaan aktiivisesti ekosysteemissä

[Kuva 2.1 Datan avaamisessa voidaan edetä vaiheittain vuorovaikutuksessa sen hyödyntäjien kanssa.]

#### 2.3.1 Datasta ilmoittaminen

Organisaation edustaja kysyy datan uudelleenkäyttöyhteisöiltä: *“Kertokaa meille, mitä tietoa haluatte ja missä muodossa, niin katsomme, mitä voimme asian hyväksi tehdä”*. Uudelleenkäyttäjät vastaavat: *“Kertokaa meille, mitä dataa teillä on, niin kerromme, mitä haluamme”*.

Aivan ensimmäinen julkaistava data tulisi olla listaus aineistoista, joita organisaatiolla ylipäätään on olemassa, vaikka nämä aineistot eivät vielä olisi saatavilla. Tällaista organisaation nettisivuilla tai laajemmassa useiden organisaatioiden tietoja yhdistävässä datakatalogissa julkaistavaa listausta kutsutaan organisaation julkisten tietoaineistojen rekisteriksi (Information Asset Registry IAR). Listaukseen sisältyvät sekä jo avoimet datavarannot että tietolähteet, jotka eivät vielä ole avoimesti saatavilla, mutta ovat kuitenkin julkisia ja joiden avoimuutta voidaan jatkossa lisätä.

Monissa organisaatioissa tällainen listaus omista tietovarannoista saattaa olla jo olemassa, toisissa sitä varten pitää tehdä omien datavarantojen inventaario. Yksi ensimmäisistä tehtävistä datan avaamista käynnistettäessä on perata olemassaolevat tietovarannot ja niihin liittyvät käytännöt.

Ensimmäinen lista organisaation datavarannoista syntyy yksinkertaisimmin muistinvaraisesti ja omia verkkosivuja selaamalla. Kattavampi listaus saadaan, jos kartoitustyöhön osallistuu useampia henkilöitä. Esimerkiksi muutaman tunnin työpajassa pystytään tuottamaan julkaisukelpoinen listaus, joka sisältää olennaisimmat tiedot organisaation hallussa olevista aineistoista. Tavoitteena ei ole vielä kattavasti selvittää kaikkia tietoja kaikista aineistoista. Alla on muutamia tärkeimpiä kysymyksistä, joihin pyritään aineistokohtaisesti vastaamaan.

- Aineiston nimi
- Lyhyt kuvaus aineiston sisällöstä
- Kuinka avointa aineisto on nykyisin, onko se saatavilla käyttöön ja jos on niin miten?

Kun julkisten tietovarantojen rekisteri on laitettu avoimesti verkkoon, voidaan siitä jo viestiä potentiaalisille datan uudelleenkäyttäjille, luoda kontakteja ja kerätä mielipiteitä. Organisaatio voi esimerkiksi jalkautua kehittäjäyhteisöihin sosiaalisen median ja tapahtumien kautta. Viestinnässä kannattaa kiinnittää huomiota kannustavuuteen ja siihen, millaisia odotuksia herättää yleisössä. Jos aineistojen avaamisesta ei ole vielä tehty päätöksiä ja halutaan pelkästään ideoita niiden hyödyntämisestä, kannattaa se kertoa.

Voi olla, ettei käyttäjien huomiota aivan alussa saavuteta. Parhaimmillaan kuitenkin jo pelkän omien datavarantojen listauksen avulla voidaan käynnistää organisaation ja datan hyödyntäjien välinen dialogi ja saada palautetta siitä, mitä dataa kannattaisi julkaista mitenkään. Lopuksi palaute ja datan avaamisesta saadut kokemukset kootaan ja arvioidaan. Mikä toimi, mikä ei? Valmistaudutaan siirtymään seuraavalle iteraatiokierrokselle, jossa pilotoidaan datan julkaisemista.

### 2.3.2 Pilottiaineistojen julkaiseminen

Kun organisaation keskeisimmät tietovarannot on tunnistettu, on tärkeää erotella huolellisesti julkaisemiseen kelpaavat aineistot muista. Tietokannoissa saattaa olla sekaisin julkisia ja salaisia tietoja. Esimerkiksi monet valtion työntekijöiden nimet ovat periaatteessa julkisia, mutta esimerkiksi hävittäjälentäjistä ei haluta kansallisen turvallisuuden vuoksi kertoa mitään. Nopean datan avaamisen kannalta tällaiset kokonaisuudet kannattaa lykätä myöhemmäksi. Erotellaan aineistot täysin julkisiin, epäselviin ja ei-julkisiin. Aluksi voidaan keskittyä pelkästään selvästi julkisten aineistojen avaamiseen.

Vielä ei kannata siivota listalta pois sellaisia aineistoja, joiden julkaisemiseen organisaatiolla ei ole oikeutta. Mahdollisuuksien mukaan myös näistä aineistoista kannattaa kertoa. Organisaation ulkopuolisten toimijoiden kannalta tällainen tieto on erittäin arvokasta hallinnon tietovarantoihin liittyvän kokonaiskuvan hahmottamisen kannalta.

Seuraavaksi suunnitellaan, ketkä toimivat organisaation yhteyshenkilöinä datan hyödyntäjiin päin ja miten esiintuleviin asioihin reagoidaan. Voidaan esimerkiksi varautua etukäteen nostamaan esiin mielenkiintoisia datan käyttötapoja ja miettiä, kuinka konfliktitilanteissa toimitaan ja miettiä vastauksia, miten esimerkiksi reagoidaan datan epämoraaliseen käyttöön.

Itse datan jakamiseen ei aluksi tarvitse tehdä minkäänlaista käyttöliittymää. Kaikkien kannalta on helpointa, jos datan hyödyntäjiä ei pyydetä rekisteröitymään tai tilaamaan aineistoa organisaatiolta, vaan julkisesti kerrotaan osoite, josta data on saatavilla. Raakadatan jakamisella pelkistetyksi tarkoitetaan, että jos aineisto on tietokantamuodossa, niin tietokannan taulujen sisältö tulostetaan vaikka pilkkueroteltuun tiedostoon, joka laitetaan internet-palvelimelle. Jos aineisto on jo valmiiksi tiedostoissa, riittää, että nämä tiedostot laitetaan julkisesti saataville.

Raakadatan jakamisen tarkoituksena ei välttämättä ole se, että joku toteuttaisi suoraan järjestelmiä siihen perustuen. Avoimesti julkaistu data auttaa kuitenkin eri tahoja perehtymään tietosisältöön ja suunnittelemaan sen hyödyntämistä. Se tekee kyseisen datan olemassaolon tunnetuksi. Kun kiinnostus tietoa kohtaan on herännyt, dataa tarjoava organisaatio voi yhdessä hyödyntäjien kanssa suunnitella sen käyttöä parhaiten palvelevat palvelurajapinnat tai muut jakeluratkaisut.

Asiaa ei ole syytä mutkistaa ja datan avausta viivyttää esimerkiksi sen takia, että organisaation resurssit eivät juuri nyt riitä rajapintojen määrittelyyn tai viimeistä piirtoa myöten hienon dataportaalin rakentamiseen. Tässäkin iterativisuus on tärkeää ja pienin askelin päästään pidemmälle. On parempi laittaa aineisto saataville ensin kevyesti, mutta kuitenkin kokonaisuutena ja rakenteellisessa muodossa, ja vasta sitten alkaa miettiä mahdollisia rajapintoja. Pinnan kiillotus ja käyttöliittymät ovat tärkeysjärjestyksessä viimeisimpiä.

### 2.3.3 Dokumentointi ja käyttötilanteet

Viitteellisen datan avaamisprosessin mukaan on tehty jo kaksi täyttä kierrosta. Ensimmäisen tuloksena julkaistiin lista kaikista organisaation hallussa olevista julkisista aineistoista. Toisen tuloksena julkaistiin ensimmäiset helposti avattavat aineistot. Samalla organisaation omien tietoresurssien tuntemus on vahvistunut ja datan käyttäjäyhteisöihin on syntynyt ensimmäiset kontaktit. Tämän seuraavan dokumentointi ja käyttötilanteet-kierroksen aikana tehdään kattavampi datavarantojen inventaario, julkaistaan uusia aineistoja ja parannetaan niiden julkista dokumentaatiota.

Inventaarioprosessi datan avaamista varten on pitkälti samantyyppinen prosessi, kuin mitä organisaatioissa saatetaan tehdä muutoinkin pelkästään sisäiseen käyttöön omien tietovarantojen jäsentämiseksi ja uudelleenorganisoidumiseksi. Inventaariossa tunnistetaan myös omia tietotarpeita organisaation työntekijöiden näkökulmasta. Tämä on pelkkää teknistä datan avaamista työläämpää, mutta toisaalta se tukee organisaation toimintaa laajemmin.

Analysoimalla käytössä olevia järjestelmiä tietohallinnon kanssa saadaan hyvä pohjakäsitys organisaation tietoresursseista ja niihin liittyvistä käytännöistä. Tietohallinnossa ei kuitenkaan välttämättä tunneta aivan kaikkia organisaatiossa käytettyjä aineistoja ja niiden käyttötapoja. Käsitystä voidaan syventää

haastatteleamalla eri käyttäjäryhmiä ja kuvailemalla käyttötilanteita, joissa organisaatiossa tarvitaan tietoa. Käyttäjien näkökulma voi auttaa löytämään uusia tiedon hyödyntämismahdollisuuksia, joita voidaan nivoa osaksi datanavausprojektia.

Tämän iteraatiokierroksen julkaisuvaiheessa voidaan kehittää jo aiemmin julkaistujen aineistojen jakelua esimerkiksi toteuttamalla avoin rajapinta (ks. luku 5.5). Aineistojen sisällön ja käyttöoikeuksien kuvailu ja tekninen dokumentaatio ovat myös tärkeitä kehityskohteita.

### 2.3.4 Tietoarkkitehtuurin kehittäminen ja sanastotyö

Pienestä datavarantojen listauksesta liikkeelle lähtenyt prosessi voi ajan myötä kulminoitua tietoarkkitehtuurin kehittämiseen, järjestelmien suunnitteluun ja organisaatorajat ylittävään tietovarantojen yhtenäistämiseen. Olennaista on, että matkan varrella avattujen datavarantojen hyödyntämisestä saattaa jo olla kertynyt kokemuksia.

Kuten aiemminkin iteraatiokierroksilla, jälleen organisaatioiden omien tietoresurssien dokumentaatiota täydennetään, aineistojen julkaisua parannetaan ja datan avaamisen oppeja otetaan mukaan strategiseen työhön. Tavoitteena on mallintaa organisaatiossa tapahtuvaa tiedon käsittelyä ja tarpeita ja kehittää tietovarantojen yhteiskäyttöä myös yli organisaatorajojen. Tämän iteraatiokierroksen julkaisuvaiheessa organisaatiolla on laadukkaita tietoaineistoja, jotka on harmonisoitu yhteiskäyttöisiksi muiden organisaatioiden kanssa ja julkaistu hyvin dokumentoidusti.

Verkkopalveluiden ja tietokantojen lisäksi kannattaa tässä vaiheessa kiinnittää huomiota myös dokumentteihin, joita käytetään toistuvasti - tällaisia voivat esimerkiksi olla taulukkolaskentatiedostot. Tiedon käyttämisen lisäksi pyritään löytämään kehittämiskohteita. Käyttäjiltä kysytään minkälaisiin tietoon liittyviin tarpeita heillä on, minkälaisiin ongelmiin he ovat törmänneet ja minkälaisia tarpeita arvelevat muilla käyttäjillä olevan. Tietolähteiden saatavuuden lisäksi selvitetään, onko tieto sellaisessa muodossa, että siitä on hyötyä vai liittyykö sen hyödyntämiseen jonkinlaista kitkaa tai mekaanista työtä.

Jos tietoarkkitehtuurista alkaa esimerkiksi löytyä dokumentteja ja kansioita, joilla ei ole selkeästi kohdennettavaa luojaa tai tekijää, tai ne on arkistoitu poikkeavasti, on tämä yleensä selkeä merkki esimerkiksi a) toimimattomasta käytännöstä, joka on ohitettu (kuten hankala toimintamalli), b) toimintatavasta, jota ei ole ohjeistettu riittävästi (ohjeiden puute) tai c) uudesta tarpeesta, jota ei ole ennakoitu (kuten uusi yksikkö, projektiryhmä etc.)

Useat nykyisistä tietoarkkitehtuurin haasteista kietoutuvat saman asian ympärille: tietoa tuotetaan erittäin runsaasti ja epäyhtenäisesti. Alkuperäisiä työvälineitä ei ole välttämättä suunniteltu aineistojen nykyisen paljouden

hallintaan. Aineistojen yhdistäminen muodostuu keskeiseksi ongelmaksi organisaatioiden sisällä ja organisaatioiden välisessä yhteistoiminnassa. Huonossa tapauksessa lähtötilanteessa ei ole saatavilla kunnollista kuvausta siitä, mitä talletettua tietoa on missäkin järjestelmässä. Kukin toimija katsoo asiasta vain omasta näkökulmastaan, tiedot ovat hajallaan eri järjestelmissä ja tietoa samoista asioista kerätään ja päivitetään useissa paikoissa.

Aineistojen kuvaamisella ja menetelmien määrittelyllä pyritään ratkomaan ja ennaltaehkäisemään yhteensopivuusongelmia ja luomaan kattava näkemys aineistoihin yli organisaatorajojen. Organisaatioiden sisällä ja myös organisaatioiden välillä pyritään pikkuhiljaa yhdenmukaistamaan tietoarkkitehtuuria ja löytämään yhteisiä tapoja kuvata tietoa.

Tietoarkkitehtuurin laatiminen edellyttää eri osapuolten organisoitua yhteistoimintaa, aikaa ja runsasta työtä. Mitä laajempi näkemys pyritään saavuttamaan, sitä työläämpää se on. Esimerkki tällaisesta tietoarkkitehtuurihankkeesta on hiljattain valmistunut kaksivuotinen KuntaGML-hanke, jonka tavoitteena oli ottaa kunnissa käyttöön standardimuotoiset rajapinnat paikkatiedolle, kuten kaavoituksen pohjatiedoille. Hyvä arkkitehtuuri palkitsee työläydestään huolimatta järjestelmien yksinkertaisempaa ylläpitona, säästyneenä työnä, tietovarantojen laadun paranemisena ja ennen kaikkea mahdollistaa tietojen laajemman uudelleenkäytön.