

Tässä dokumentissa ideoidaan, minkälainen olisi minimalistinen opas, jolla organisaatio pääsee datanavauksessa alkuun. Ideana on kuvata datanavaukseen liittyen eri askelmia ja määritellä, mitä vaaditaan, että päästään ko. askeleelle.

Osittain samoja näkökulmia löytyy seuraavien linkkien takaa. Niistä voisi poimia keskeisiä sisältöjä tänne...:

- <http://muistio.tieke.fi/avoindata-vaatimukset> (järjestelmävaatimuksia avoindata huomioiden)
- <http://rinnetmaki.net/howto/avoindata.html> (Samuel Rinnetmäen tekemä muistio)
- <https://gist.github.com/1548127> (Julkinen Data -oppaan 5-luku: Tekninen valmistautuminen)

Data-avaajan taso 1: Kertaluontoinen aineistojulkaisu

Kuvaus: Organisaatiolla on aineistoa, joka tarjotaan käyttöön sellaisenaan. Avaus on kertaluontoinen aineiston julkaisu. Tiedosto voi olla esim. excel-tiedosto, csv-tiedosto tai muu ascii-tiedosto.

Kertaluontoisessa julkaisussa ei tarvitse kiinnittää huomiota jatkuvuuteen tiedostosisällön kannalta eikä tarvitse sitoutua esim. tiettyihin päivitysväleihin.

Vaatimukset:

- Määritellään aineistolle lisenssi: XXX -- Tähän linkki listaan, minkälaisia lisenssejä löytyy ja mitä hyvää/huonoa kussakin on
- Laitetaan aineisto ladattavaksi siten, että se on tietyllä nimellä aina ladattavissa. Latauspaikka voi olla esim.
 - Organisaation oma webbipalvelu. Kannattaa tehdä aineistojulkaisuille oma virtuaalihakemisto, jotta latauslinkki on lyhyt ja on helpompi taata, että linkki pysyy samana, vaikka julkaisujärjestelmä vaihtuu. Esim. <http://organisaatio.fi/download/AsuntojenMyyntihinnat2010.xls>
 - Google Docs -dokumentti (esim. Excel-tiedoston voi ladata Google Spreadsheet-taulukoksi ja antaa siihen kaikille lukuoikeus)
- Tehdään aineistokuvaus (mitä aineistossa on).
 - Mitä aineistosta pitäisi kuvailla (minimissään):
 - Milloin aineisto on irrotettu
 - Aineiston muuttujien ja käytettyjen standardien kuvaus
 - Yhteystiedot tai yhteydenottomahdollisuus (vai pitäisikö olla vapaaehtoinen)
 - Aineistokuvauksen sijainti voi olla esim.
 - Organisaation oma webbipalvelu
 - Google Docs -dokumentti, johon avataan vain-luku-tyyppinen oikeus kaikille (ei saa vaatia Google-tunnuksia dokumentin avaamiseen)
 - Erikseen perustettava blogi (esim. Posterous, Google Blogspot tns.)
 - Datakatalogin wiki <mihin datakatalogiin ohjeistuksessa olisi hyvä

viitata>?

- Julkaistaan tieto aineistosta
 - <tähän tieto, mihin datakatalogiin tms. tiedot aineistosta + lisenssistä voi laittaa>
- Valinnainen: Ohjeistetaan, mihin datan käyttäjien tulee ilmoittaa datankäyttö
 - Esim. Google Forms -lomake, johon voi syöttää datan käyttäjän email-osoitteen ja kuvauskenttä, johon voi kertoa, missä dataa käytetään
 - Tälle listalle voidaan lähettää tiedotteita dataan liittyen
- Valinnainen: Ohjeet статистиikan keräämiseen aineiston latauksista

Data-avaajan taso 2: Toistuva aineistojulkaisu

Kuvaus: Organisaatiolla on aineistoa, joka tarjotaan käyttöön sellaisenaan. Avaus tehdään niin, että tiedostonimet sisältävät aikajakson ja julkaisu on jatkuvaa (esim. kerran/kk, 4 kertaa vuodessa jne). Toistuvassa julkaisussa myös tiedon formaatin tulisi pysyä samana julkaisusta toiseen. Esimerkiksi julkaisun ollessa esim. Excel-tiedosto, tulisi sarakkeiden jne. pysyä samana.

Vaatimukset ovat samat kuin tasossa 1. Sen lisäksi tulee seuraavat uudet vaatimukset:

- Tiedoston nimeäminen standardoidaan ja ohjeistetaan. Lataustiedoston nimi voi olla esim. <http://organisaatio.fi/download/AsuntojenMyyntihinnat2010.xls>, mutta ohjeistuksessa kerrotaan, että nimiosan viimeiset 4 numeroa kuvaavat vuoden. Näin tiedon käyttäjä voi olettaa, että vuoden päästä on ladattavissa uusi aineisto
- Tiedoston sisällön formaatti pysyy julkaisusta toiseen samana. Mikäli formaatti muuttuu, siitä tiedotetaan datan käyttäjiksi rekisteröityneille ja päivitetään ohjeisiin tieto muutoksesta

Data-avaajan taso 3: Aineiston reaaliaikainen julkaisu

Kuvaus: Tasolla 3 organisaatio tarjoaa esim. REST-rajapintaa, jolla voidaan tehdä reaaliaikaisia aineistohakuja. Kun aiemmilla tasoilla julkaistiin staattisia dokumentteja, tässä avaustasossa tuotetaan datasisältö palvelimella reaaliaikaisesti.

Vaatimukset ovat samat kuin tasoilla 1 ja 2. Sen lisäksi:

- Toteutetaan REST...
- ...tähän tulisi nyt esimerkkikuvaukset, miten tehdään simppeli REST-palvelu..

Yksinkertaisen REST-rajapinnan sarakemuotoiselle CSV-datalle voi toteuttaa avoimen lähdekoodin Webstore-ohjelmistolla (<https://github.com/okfn/webstore>). Data tallennetaan SQL-tietokantaan (myös sqlite käy). Webstore on toteutettu Python-ohjelmointikielellä. Katsotaan josko suomalaisen fi.thedatahub.org CKAN-instanssiin saataisiin mukaan mahdollisuus datan tallentamiselle myös webstore-muodossa. Tämä toiminto voidaan toteuttaa CKAN:in Webstore-laajennuksella (<http://wiki.ckan.org/Webstore>).

