

soRvi avoimen datan työkalupakki

Leo Lahti*, Juuso Parkkinen[†] ja Joona Lehtomäki[‡]

2. helmikuuta 2012

1 Johdanto

Kaikille avointen yhteiskunnallisten tietovarantojen määrä on voimakkaassa kasvussa [2]. Talouteen, säähän, paikkatietoon, liikenteeseen, koulutukseen ja muihin alueisiin liittyvää dataa on alettu avaamaan julkishallinnon toimesta Suomessa ja muualla. Aineistojen pöyhintä avaa uusia näkökulmia ja avoimuus mahdollistaa yhdistämisen toisiin tietolähteisiin, jolloin voidaan vastata kysymyksiin joihin yksittäisillä aineistoilla ei päästä käsiksi. Laaja saatavuus ja käyttö voi siten merkittävästi nostaa datan arvoa.

Laskennallisten työkalujen saatavuus on osoittautunut keskeiseksi pullonkaulaksi. Tämä projekti pyrkii paikkaamaan puutetta tarjoamalla yleiskäyttöisiä välineitä julkisten tietoaaineistojen hakuun, putsamiseen, yhdistelyyn, louhintaan ja visualisointiin. Pakettiin on koottu kattava kokoelma ratkaisuja avoimen datan käsittelyyn. Myös uudet ehdotukset ja lisäykset ovat tervetulleita.

Yksityiskohtaisempia esimerkkejä paketin käytöstä Suomi-datan penkomiseen löydät Louhos-blogista¹. Ylläpitäjien yhteystiedot löytyvät projektin kotisivulta kautta².

2 Paketin asennus

Yksityiskohtaiset asennusohjeet löytyvät osoitteesta
<http://sorvi.r-forge.r-project.org/asennus.html>

*Wageningen University, Nederland <leo.lahti@iki.fi>

[†]Aalto-yliopisto <juuso.parkkinen@gmail.com>

[‡]Helsingin yliopisto

¹<http://louhos.wordpress.com>

²<http://sorvi.r-forge.r-project.org>

3 Pakettiin viittaaminen

Paketin tarjoamat välineet ovat vapaasti käytettävissä FreeBSD-lisenssillä³. Mikäli paketista on apua, toivomme viittaamista työhön [1]. Lisää tietoa löytyy projektin kotisivulta⁴.

4 Julkiset tietokannat

Alla esimerkkejä joidenkin julkisten aineistojen lataamisesta sorviin.

4.1 Helsingin seudun ympäristöpalvelut

Helsingin seudun ympäristöpalveluiden⁵ avoimia aineistoja ((C) HSY 2011) voi hakea valmiilla hakurutiinilla. Seuraava esimerkki lataa väestötason tietoja ja kuvaa ne Helsingin kartalla. Lisätietoja aineistoista löytyy HSYn verkkosivuilta⁶.

```
> library(sorvi)
> sp <- GetHSY("Vaestoruudukko")
> df <- as.data.frame(sp) # investigate data contents
> at <- c(seq(0, 2000, 250), Inf)
> q <- PlotShape(sp, "ASUKKAITA", type = "oneway", at = at, ncol = length(at))
```

4.2 Maanmittauslaitos

Suomen karttatietoja on haettu Maanmittauslaitoksen sivuilta⁷ rajapintasyistä valmiiksi sorviin⁸. Seuraava komento lataa valmiin listan MML:n shape-muotoisia aineistoja ja poimii kuntarajat 1:1000000 kartalta:

```
> data(MML)
> sp <- MML[["1_milj_Shape_etr_shape"]][["kunta1_p"]]
```

Alkuperäisten shape-tiedostojen suora luku ja esikasittely onnistuu seuraavasti:

```
> sp <- readShapePoly("kunta1_p.shp")
> sp <- PreprocessShapeMML(sp)
```

³http://en.wikipedia.org/wiki/BSD_licenses

⁴<http://sorvi.r-forge.r-project.org/>

⁵<http://www.hsy.fi/seututieto/kaupunki/paikkatiedot/Sivut/Avoindata.aspx>

⁶http://www.hsy.fi/seututieto/Documents/Paikkatiedot/Tietokuvaukset_kaikki.pdf

⁷<http://www.maanmittauslaitos.fi/aineistot-palvelut/digitaaliset-tuotteet/ilmaiset-aineistot/hankinta>

⁸Jatkokaytto sallittu - (C) MML 2011 - <http://www.maanmittauslaitos.fi/node/6417>

4.3 Tilastokeskus

Tilastokeskuksen sivuilta⁹ löytyy avoimia aineistoja PC-Axis muodossa. Tässä esimerkkinä poimittu kuntien mediaanitulot (määrittele muuttujaan url-osoite ”http://pxweb2.stat.fi/Database/StatFin/tul/tvt/2009/120_tvt_2009_2011-02-18_tau_112.fi.px”:

```
> library(pxR)
> px <- GetPXTilastokeskus(url)
```

4.4 Suomen ympäristökeskus

Suomen ympäristökeskuksen¹⁰ (SYKE) OIVA-palvelu¹¹ sisältää lukuisia avoimia aineistoja. Aineistot ovat saatavilla sekä latauspalvelun että WMS-rajapinnan kautta. OIVA-palvelu vaatii ilmaisen käyttäjätunnuksen, WMS-palvelu ei edellytä tunnusta. Esimerkkejä näiden aineistojen käytöstä löytyy Louhos-blogista.

4.5 Muuta

4.5.1 Postinumerot

```
> suomen.postinumerot <- GetPostalCodeInfo()
```

4.5.2 Kuntatason informaatio

Poimi kuntatason tietoja verkosta ja muodosta taulukko:

```
> municipality.info <- GetMunicipalityInfo()
```

Etsi annetuille kunnille vastaava maakunta:

```
> kunnat <- as.character(municipality.info$Kunta)
> kunta.maakunta <- FindProvince(kunnat, municipality.info)
```

4.5.3 Maakuntatason informaatio

Poimi maakuntatason väestötiheystiedot Wikipediasta:

```
> url <- "http://fi.wikipedia.org/wiki/V%C3%A4est%C3%B6tiheys"
> vaestotiheys <- GetProvinceInfo(url)
```

⁹<http://www.stat.fi/tup/tilastotietokannat/index.html>

¹⁰<http://www.ymparisto.fi/syke>

¹¹www.ymparisto.fi/oiva

5 Versiotiedot

```
> sessionInfo()
```

```
R version 2.14.1 (2011-12-22)
```

```
Platform: x86_64-pc-linux-gnu (64-bit)
```

```
locale:
```

```
[1] LC_CTYPE=en_GB.UTF-8      LC_NUMERIC=C
[3] LC_TIME=en_GB.UTF-8      LC_COLLATE=en_GB.UTF-8
[5] LC_MONETARY=en_GB.UTF-8  LC_MESSAGES=en_GB.UTF-8
[7] LC_PAPER=C               LC_NAME=C
[9] LC_ADDRESS=C            LC_TELEPHONE=C
[11] LC_MEASUREMENT=en_GB.UTF-8 LC_IDENTIFICATION=C
```

```
attached base packages:
```

```
[1] stats      graphics  grDevices  utils      datasets  methods    base
```

```
loaded via a namespace (and not attached):
```

```
[1] tools_2.14.1
```

Viitteet

- [1] Leo Lahti, Juuso Parkkinen ja Joona Lehtomäki (2011). soRvi - avoimen datan työkalupaketti URL: <http://louhos.wordpress.com>
- [2] Antti Poikola, Petri Kola ja Kari A. Hintikka (2010). Julkinen data – johdatus tietovarantojen avaamiseen Liikenne- ja Viestintäministeriö. Edita Prima Oy, Helsinki 2010. URL: <http://www.julkinendata.fi>