

A stylized topographic map of Finland, rendered in purple lines, occupies the left side of the slide. The map shows the country's outline and internal contour lines representing elevation.

KÄYTÄNNÖN ESIMERKEILLÄ KIINNI DATAAN

TOPI TJUKANOV
GISPO OY

GISPO



KOULUTUSPÄIVÄT

Tiistai 21.5.2019: Dataa monessa muodossa

Tiistai 28.5.2019: Käytännön esimerkeillä kiinni dataan

Tiistai 4.6.2019: Visualisoidaan tietoa

Keskiviikko 12.6.2019: Tieto liiketoiminnassa



KOTITEHTÄVÄT

- Jokaisen koulutuspäivän jälkeen on mahdollisuus tutustua lisätietopakettiin ja vastata muutamaiin koulutusta tukeviin kysymyksiin verkkokoulutusalan kautta
<https://gispolearning.thinkific.com/courses/datan-hyodyntaminen-hamk-koulutus/>
- Kaikki koulutuspäivien materiaalit lisätään samaan paikkaan, eli jos et pääse osallistumaan lähiopetuspäivään, materiaalit on kuitenkin saatavilla jälkikäteenkin pari kuukautta. Koulutettaville on myös välitetty materiaalit pdf:nä.

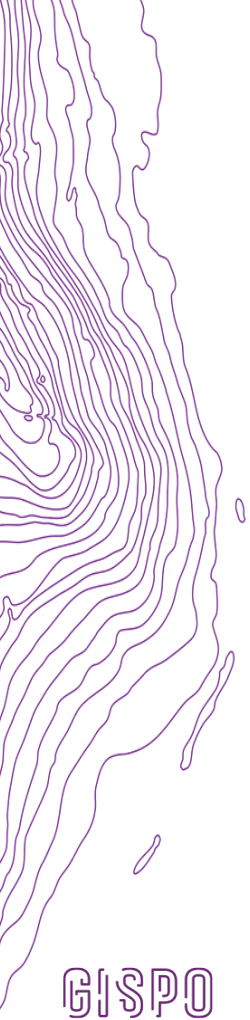


PÄIVÄN OHJELMA

- Tietolähteet ja tiedon löytäminen
 - Suomalaiset aineistot
 - Kansainväliset aineistot

LOUNAS

- Rajapintojen perusteet
 - REST
 - Paikkatietorajapinnat
- Tietokantojen perusteet
 - Yleistä tietokannoista
 - SQL



ESITTELYT

GISPO OY

Perustettu 2012

Avoimuus voittaa!

Koulutamme avoimen datan ja avoimen lähdekoodin paikkatieto-ohjelmistojen käyttöön

www.gispo.fi

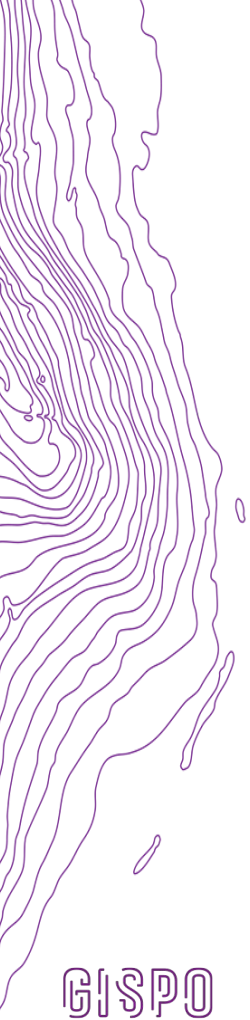
info@gispo.fi

<https://www.facebook.com/GispoFinland/>

Topi Tjukanov - paikkatietokonsultti & osakas

GISPO Twitter: @tjukanov





TIETOLÄHTEET JA TIEDON LÖYTÄMINEN

GISPO



You've gone incognito

Now you can browse privately, and other people who use this device won't see your activity. However, downloads and bookmarks will be saved. [Learn more](#)

Chrome **won't save** the following information:

- Your browsing history
- Cookies and site data
- Information entered in forms

Your activity **might still be visible** to:

- Websites you visit
- Your employer or school
- Your internet service provider



MISTÄ LÖYTÄÄ AVOINTA DATAA

- Tilastokeskus: <https://www.stat.fi/index.html>
- Maanmittauslaitos:
<https://www.maanmittauslaitos.fi/asioi-verkossa/avoimien-aineistojen-tiedostopalvelu>
- THL: <https://sotkanet.fi/sotkanet/fi/index>
- Ja monet muut julkiset organisaatiot!

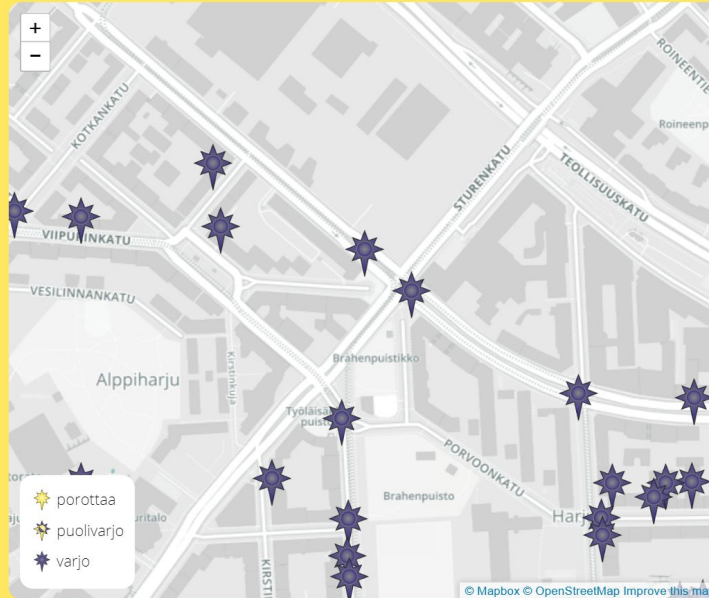
AINEISTOA VOI SAADA PYYTÄMÄLLÄ



POROTTA - TERASSIBONGARI



Helsinki
23 °C



HAE TERASSIA KALLIOSSA:

LÄHIMMÄT ▾

Ei terasseja.

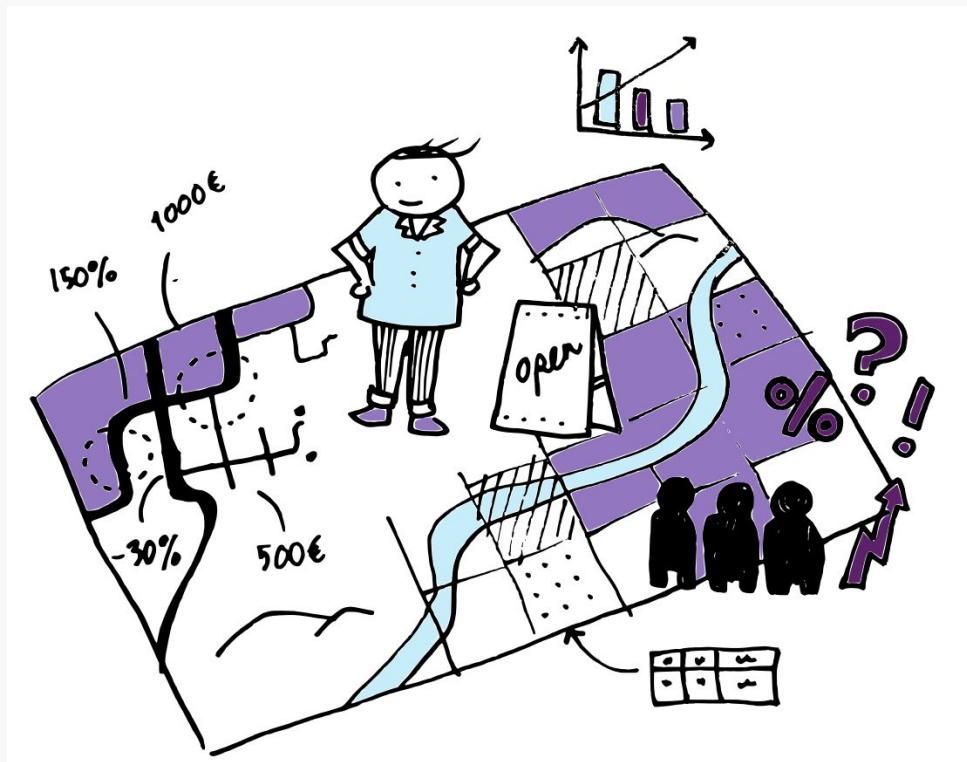
Tietoa palvelusta

Palaute

SUPERGISSAAJAT 2013

MISTÄ TE LÖYDÄTTE TIETOA OMASSA TYÖSSÄNNE?

- Keskustellaan hetki aiheesta
- Mistä tietoa löytyy?
- Mitä rekistereitä / tiedontuottajia käytetään työssä?



YRITYSREKISTERIT

Yritystietoja keräävät rekisterit

- Yritystietorekisteri eli YTJ
- Patentti- ja rekisterihallitus eli PRH

Muita tiedon kerääjiä:

- Tilastokeskus
- Verohallinto
- Kaupalliset toimijat: Suomen Asiakastieto, Fonecta, Posti, xx?
- Muita?

GISPO

Yritystiedot

Virallinen nimi	Gispo Oy Lähde: YTJ
Rinnakkais- toiminimet	Gispo Ltd Lähde: YTJ
Yhtiömuoto	Osakeyhtiö Lähde: YTJ
Perustamisvuosi	2012 Lähde: Fonecta-konserni
Kotipaikka	Espoo Lähde: YTJ
Yrityksen liikevaihtoluokka	Ei tiedossa Lähde: Fonecta-konserni
Toimipaikan henkilöstöluokka	1–4 henkilöä Lähde: Fonecta-konserni
Toimialat	IT-konsultointia, IT-palveluja (päätoimiala) Lähde: Fonecta-konserni
Toimialaluokitus (TOL2008)	6202 Atk-laitteisto- ja ohjelmistokonsultointi Lähde: YTJ
Voimassa olevat rekisteröinnit	Ennakkoperintärekisteri: 19.1.2012 (tarkistus 28.2.2018) Verohallinnon perustiedot: 31.1.2012 Työnantajarekisteri: 1.3.2015 Kaupparekisteri: 30.1.2012 Arvonlisäverovelvollisuus: Liiketoiminnasta: 17.1.2012 Lähde: YTJ

Kuvankaappaus: Fonecta, Finder.fi

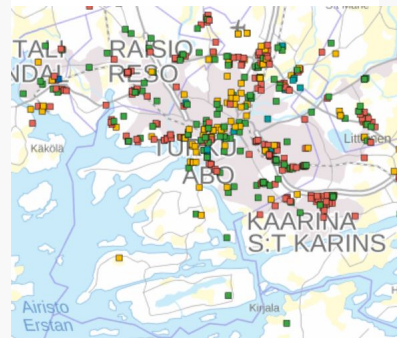
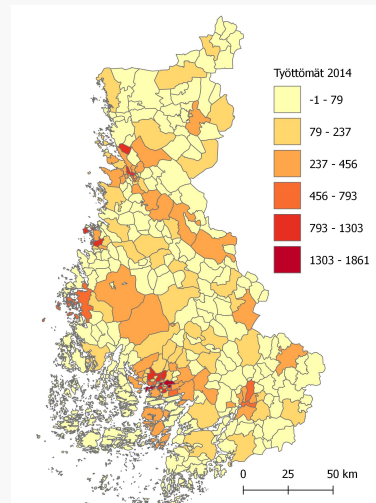
TYÖLLISYYSTILASTOT JA SIJAINTI

Työpaikkojen ja työttömien määristä on saatavilla jonkin verran tietoa avoimena datana, mutta joskus jopa parin tai neljän vuoden viiveellä ja yleensä ilman tarkkaa sijaintitietoa. Myös muutamat kaupungit ovat alkaneet tarjoamaan työpaikkatietoa avoimena datana.

Sijaintitiedon tarkkuus vaikuttaa tarkasteluun. Sijaintitieto eli paikkatieto on erityisesti silloin tarpeen, jos halutaan tarkastella tietoja myös kartalla. Sijainti voidaan lisätä tietoihin alueena (kunta, maakunta, postinumeroalue) tai osoitteena (toimipaikka, kotiosoite) tai tarkkana X,Y-koordinaattina.

Paikkatietoja voidaan tarjota tiedostomuodossa tai rajapintana, molemmat vaativat paikkatieto-ohjelmiston / asiakassovelluksen joka osaa lukea näitä formaatteja

GISPO



Kartta-aineistojen lähteet: Tilastokeskus (Paavo-postinumeroaineisto & Tuotanto- ja teollisuuslaitokset)



TILASTOKESKUKSEN TYÖLLISYYTEEN JA YRITYKSIIN LIITTYVÄT AINEISTOT JA REKISTERIT – AVOIMET JA MAKSUTTOMAT TIEDOT

- Mielenkiintoisia mm. **Paavo postinumeroaineisto** sekä **StatFin** -tilastotietokanta
 - Paavo: <http://www.stat.fi/tup/paavo/index.html>
 - Tilastot: <http://www.stat.fi/org/avoindata/pxweb.html>
 - Lisäksi esim. Tuotanto- ja teollisuuslaitokset saatavina pisteinä: http://www.stat.fi/tup/rajapintapalvelut/inspire_aineistot.html
- Päivitystiheys: usein vuosittain tai joskus kvartaaleittain
- Avoimissa dataista suhteellisen karkea tilastointiyksikkö:
 - Aloittaneiden ja lopettaneiden yritysten määrä / **kunta**
 - **Maakunnittain** saadaan tietoa toimipaikkojen lukumäärästä, liikevaihdosta ja henkilöstömäärästä
 - Avoimet työpaikat toimialan mukaan **suuraluettain**
- Lounaistiedossa myös mukavasti visualisoitu alueen työllisyystietoja tilastokeskuksen tietojen avulla:
<http://www.lounaistieto.fi/tietopalvelut/tilastot/tyopaikat-ja-tyollisyys/>



TILASTOKESKUKSEN AINEISTOT JA REKISTERIT – MAKSULLISET/LUVANVARAISET AINEISTOT

- **YKR-ruutuaineistosta** (250m x 250m, osin 125m x 125m)
 - Liiteri-palvelun kautta, jos tähän on tunnukset (voi anoa käyttöoikeutta SYKE:ltä).
 - Työvoima toimialoittain
 - Työpaikat
 - Työmatkan pituus asuinpaikan ja työpaikan mukaan toimialoittain
- **Yritysrekisteri:** <http://www.stat.fi/tup/yritysrekisteri/tietoseloste.html>
 - ⇒ tietoa myös liikevaihdosta, henkilöstömäärästä
 - yksittäisiä tietoja Y-tunnuksen avulla yrityshakupalvelusta: <http://www.stat.fi/yty/> (ei anna tuloksena esim. liikevaihtoa)



KANSAINVÄLISET TIETOLÄHTEET

- EU Open Data Portal: <https://data.europa.eu/euodp/en/home>
- World Bank: <https://data.worldbank.org/>
- Huge list of API's: <https://github.com/toddmotto/public-apis>
- Free GIS data: <https://freegisdata.rtwilson.com/>
- OpenStreetMap: <https://www.openstreetmap.org/>



OPEN DATA CAN ANSWER YOUR QUESTION

**IF YOU DEFINE THE QUESTION
AFTER YOU HAVE FOUND THE DATA**



KÄSITTEISTÖÄ

- Metadata = dataa datasta
- Datan skreippaus = datan kerääminen “väkivallalla”
- Dataportaali = sivusto, jonne on kerätty dataa

Helsingissä on jo lähes 2 700 Airbnb-asuntoa – katso kartalta, missä kohteita on eniten ja paljonko yöpyminen eri alueilla maksaa

Vaikka Airbnb-toiminta ammattimaistuu, ei kotimajoituspalvelu vielä uhkaa Helsingin hotellien toimintaa. Airbnb-bisnekselle vaaditaan silti samanlaisia sääntöjä kuin majoituslalla.



Liisa Niemi HS Juhani Saarinen HS Topi

Tjukanov

Julkaistu: 16.6.2018 2:00 , Päivitetty: 16.6.2018

6:44

MATKAILUN KASVU Helsingissä näkyy myös vuokrapalvelu Airbnb:n majoituskohteiden määrässä. Yhä ammattimaisemmaksi muuttuva



Luetuimmat

JUURI NYT

PÄIVÄ

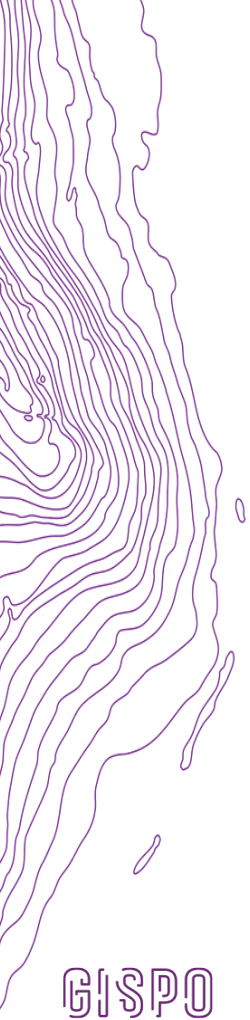
VIIKKO

1. Emmi-Liia Sjöholm tiesi sen olevan virhe, mutta silti hän asettui laverille, raotti paitaansa ja antoi tatuoijan upottaa neulan kylkeensä – nyt hän on maksanut virheestään tuhansia euroja [Tilaa jille](#)
2. Kaksi pyöräilijää loukkaantui sairaalakuuntoon törmättyään yllättäen esteeseen Töölössä
3. Mika Lehtosuo saa potkut HJK:sta, kertoo lähde HS:lle – seura kertoo uudesta päävalmentajasta, suora lähetys kello 11.55
4. **HS Helsinki:** Jos tämä myyjä ei uskalla vitsailla sinulle, on ehkä syytä olla huolissaan: Hämmäntävää läppää heittävä Dadroei Hassan nousi ilmiöksi Vuosaarella
5. **HS Vantaa:** Äiti kulki kaksivuotiaan lapsensa perässä Hoplopin leikkivälillä Vantaalla ja järkyttyi: "Likaa ei voinut olla huomaamatta"
6. "Ootko sä jutsku?" Itäkeskuksessa eurovaaliedokkaalle esitetty kysymys alleviivaa Euroopassa elettyjä vakavia aikoja



HARJOITUS 1: TIEDON LÖYTÄMINEN

- Tutkitaan avoimen datan portaaleja:
 - Avoindata.fi
 - Hri.fi
 - Tilastoksekus
 - EU Open Data Portal
- Millainen data on suosituinta?
- Etsi tietyn datan metadatat: mitä ne kertovat?
- Mieti joku käyttötapaus johon voisit tarvita dataa ja kokeile löydätkö siihen tarkoitukseen sopivaa dataa



RAJAPINTOJEN PERUSTEET

GISPO



RAJAPINTA?

- Ohjelmointirajapinta (eng. API = Application Programming Interface)
- Rajapinta tarjoaa reaaliaikaisen pääsyn tietoon tai toiminnallisuuteen
- Järjestelmäriippumaton lähestymistapa
- Rajapinta voi olla avoin tai suljettu



Swiss sockets



Type 12 triple socket (10 A), no longer sold or installed



Type 13 triple socket (10 A)



Very outdated Type 14 socket



Type 15 socket (3-phase, 10 A)



Type 23 triple socket (16 A)



Type 25 socket (3-phase, 16 A)



TIEDOSTOPOHJAISUUS VS. REAALIAIKAINEN RAJAPINTA

- Miksi on järkevämpää käyttää tietoja suoraan rajapinnasta kuin tiedostosta?
- Saat vain sen minkä haluat
 - Tiedon määrä hallittavissa
- Tieto päivittyy itsestään
 - Reaaliaikaisuus



RAJAPINTOJEN KÄYTTÖESIMERKKEJÄ

- Liikennöintiyhtiö tarjoaa omiin aikatauluihinsa avoimen rajapinnan, jotta reitityssovellusten kehittäjät voivat rakentaa omia reitityssovelluksiaan
- Kunta haluaa kerätä palautetta tien kunnosta kuntalaisilta, joten se avaa palauterajapinnan
- Keksitkö muita esimerkkejä?



RAJAPINTAESIMERKKEJÄ

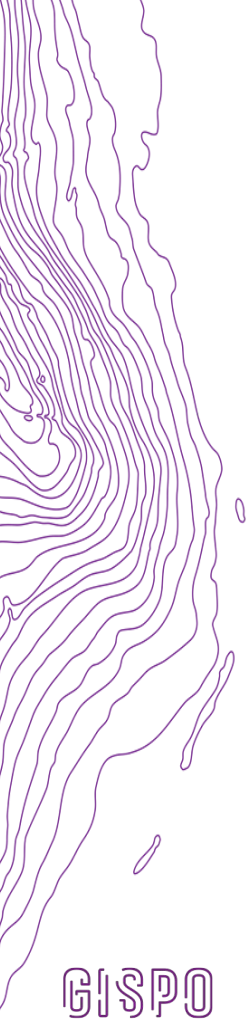
Palauterajapinta tarjoaa helpon kanavan antaa palautetta kaupungille ja seurata palautteen etenemistä.

Tapahtumarajapinta (Linked events) helpottaa tapahtumien löytämistä tarjoamalla tapahtumatiedot yhtenäisessä formaatissa.

Päätösrajapinta lisää kaupungin päätösten näkyvyyttä tarjoamalla datan rakenteisessa ja yhdenmukaisessa muodossa.

Resurssienvarausrajapinta tuo teknisen valmiuden varata kaupungin tiloja digitaalisesti.

Lähde: 6Aika avoimet rajapinnat



GISPO

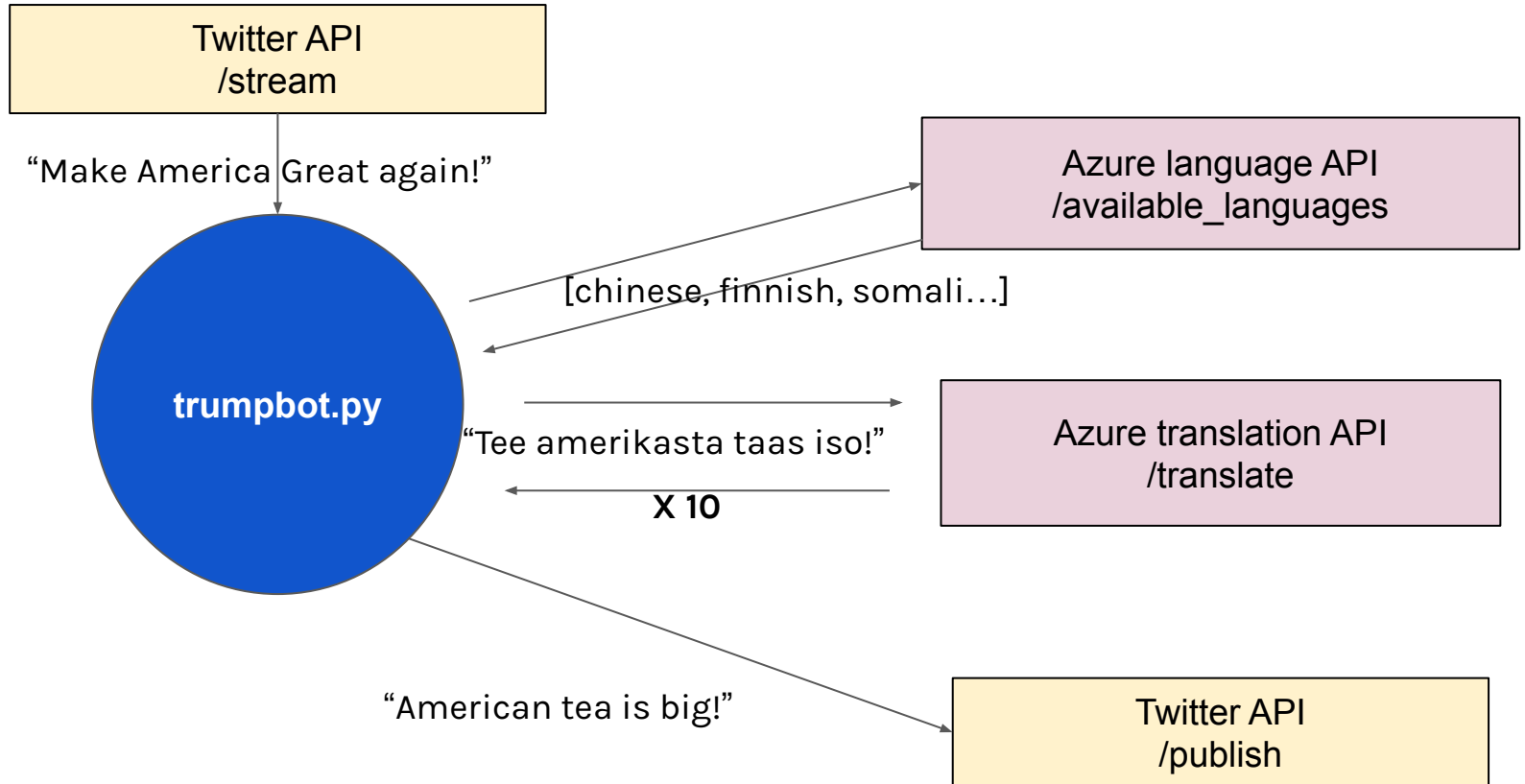


Lost in Trumplation

@lost_trump

Chinese whispers. I am a robot that translates [@realDonaldTrump](#) tweets through random languages and back to English. Created by [@tjukanov](#)

TWITTERBOTIN “ARKKITEHTUURI”



PRH:N AVOIMEN DATAN RAJAPINTA

- <http://avoindata.prh.fi/>
- YTJ:n ja PRH:n ohjelmointirajapinnat
- Tietoja saatavilla seuraavista: osakeyhtiö, julkinen osakeyhtiö, asunto-osakeyhtiö, osuuskunta, vakuutusyhtiö ja julkinen vakuutusyhtiö.
- Esimerkkihaun voi tehdä sivuilla y-tunnuksen avulla (testataan)
- Jos ei halua opetella koodausta, onneksi datasta löytyy ns. Datadumppi avoindata.fi:n kautta - aineisto on .csv-muodossa
- Yksittäisiä tietoja voi toki etsiä <https://virre.prh.fi/> tai <https://www.ytj.fi/index.html> kautta



PROSESSI? RAJAPINNALTA DATAA KÄYTTÖÖN

- Esimerkkinä <https://www.digitraffic.fi/rautatieliikenne/>
 1. Kun sopiva data on löytynyt, tutkitaan dokumentaatiota
 2. Otetaan data käyttöön parhaaksi soveltuvalla tavalla
 - a. Koodi? Käyttösovellus? Upotus verkkosivuille?
 3. Tarkastetaan prosessin toimivuus ja iteroidaan



REST API

- Representational State Transfer
- HTTP-protokollaan perustuva malli rajapintojen toteuttamiseen
- Toiminnot GET, HEAD, POST, PUT, PATCH, DELETE, CONNECT, OPTIONS ja TRACE
- Esimerkki REST rajapinnasta:
<https://developer.here.com/documentation>
- GraphQL, MQTT...



PAIKKATIETORAJAPINNAT WMS & WFS

- WMS = karttakuvapalvelu
- WFS = vektoriaineistoja (tiedostoja) palauttava rajapintapalvelu
- <https://gis pohelp.zendesk.com/hc/fi/articles/208159815-Avoimen-datan-WMS-ja-WFS-karttapalveluita>
- Esimerkkidemo QGIS:ssä



OpenRefine

A free, open source,
powerful tool for working
with messy data



Home

Community

Documentation

Download

Contact Us

Blog

Welcome!

OpenRefine (formerly Google Refine) is a powerful tool for working with messy data: cleaning it; transforming it from one format into another; and extending it with web services and external data.

OpenRefine is available in English, Chinese, Spanish, French, Russian, Portuguese (Brazil), German, Japanese, Italian, Hungarian, Hebrew, Filipino, Cebuano, Tagalog

OpenRefine is supported by:

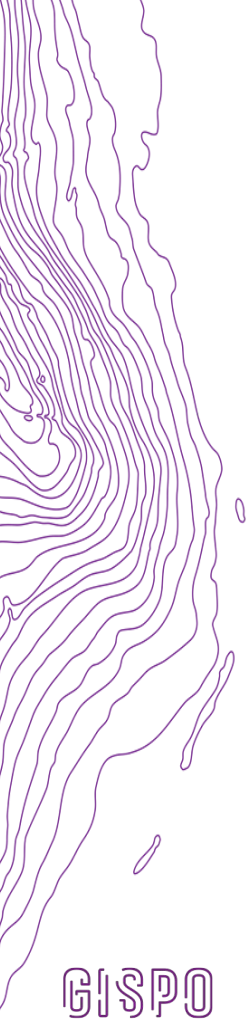
 **Google News Initiative**

Introduction to OpenRefine

A purple line-art topographic map of a coastline, showing contour lines and a few islands, positioned on the left side of the slide.

HARJOITUS 2: DATAA RAJAPINNALTA

- Kokeillaan miten dataa voi saada ulos rajapinnalta
- Miten API:t toimivat?
- Miten voin käsitellä aineistoja OpenRefinellä?



TIETOKANTOJEN PERUSTEET

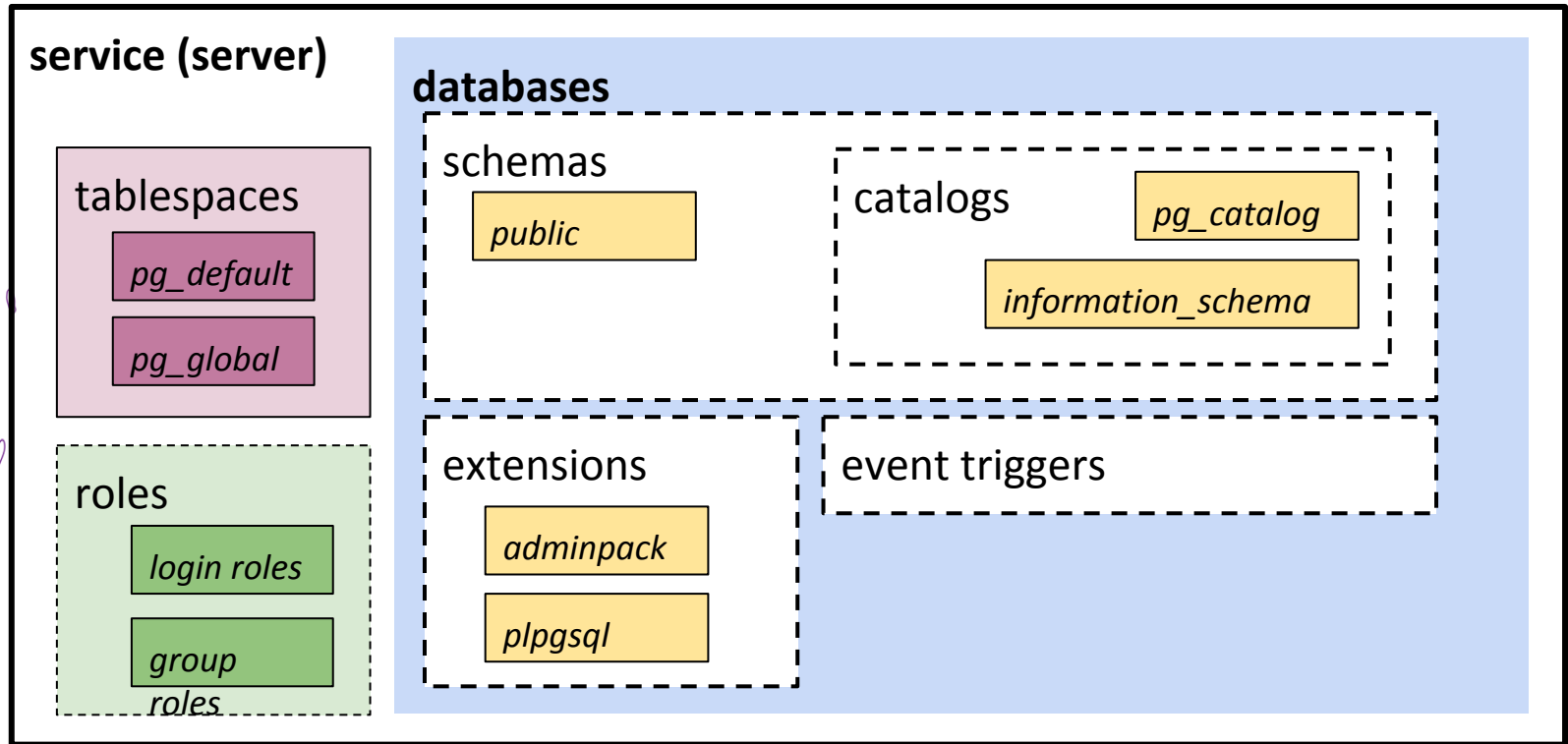
GISPO

A decorative topographic map of Finland is shown on the left side of the slide, rendered in purple lines. The map shows the coastline and internal topography with contour lines.

YLEISTÄ TIETOKANNOISTA

- Tietokanta = tietokoneen kovalevyllä sijaitseva tietojen kooste
- Tietokantaohjelmistoja esim. PostgreSQL, Oracle DB, MySQL, MongoDB...
- Tietokanta on lähes aina sovellusten taustalla, mutta sitä voi käyttää myös puhtaasti datan analysointiin
- Kanta voi sijaita omalla koneella tai serverillä jossain muualla

POSTGRESQL RAKENTEITA



A purple line-art topographic map of Finland is positioned on the left side of the slide, showing contour lines and the outline of the country.

SQL

- Structured Query Language
- Yleinen tietokantojen kysely/skriptauskieli
- NoSQL = Not Only SQL
- SELECT, UPDATE, INSERT DELETE



SQL ESIMERKKEJÄ

```
SELECT * FROM mydatabase
```

```
SELECT * FROM mydatabase WHERE value = 1 AND name LIKE 'A%'
```

```
DELETE FROM mydatabase where value is NULL
```



MILLOIN HYÖDYN TIETOKANNASTA?

- Suuret datamassat tai nopeasti päivittyviä tietoja
- Yhtenäinen tietomalli
- Datan yhdistäminen ja relaatiot
- Rakennat sovellusta
- “Kun Excel ei riitä”
 - Tiedon hakeminen useammalla parametrilla
- Käytännön esimerkkinä esimerkiksi asiakasrekisteri

ESIMERKKI DATA-ANALYYSISTÄ TIETOKANNAN AVULLA

Searching for isolation with GIS



Topi Tjukanov

Jul 10, 2018 · 8 min read

Searching for the nearest neighbor is a classic GIS question. What is the nearest bar from my location? How close is the nearest entrance to the metro from this road? When comparing the distance between two objects (source and target), with modern GIS tools it is relatively simple to calculate the distances and sort the results.

What if you first have to find the target point and just have a source point and an empty plane? Obviously the closest target is at the same location as the source, but what about vice versa?

DEMO DATAN HAKEMISESTA TIETOKANNASTA

The screenshot shows the QGIS DB Manager window. The left sidebar contains a tree view of providers, with 'postgres11' selected under the 'PostGIS' provider. The main panel displays the connection details for 'postgres11'.

postgres11

Connection details

Host: localhost
User: postgres
Database: postgres

General info

Server version: PostgreSQL 11.2, compiled by Visual C++ build 1914, 64-bit

PostGIS

Library: 2.5.1
GEOS: 3.7.0-CAPI-1.11.0 3.7.1
Proj: Rel. 4.9.3, 15 August 2016

Privileges

User has privileges:

- create new schemas
- create temporary tables

The QGIS interface also shows a top menu bar (Project, Edit, View, Layer, Settings, Plugins, Vector, Raster, Database, Web, Mesh, Processing, Help), a toolbar, and a bottom status bar with coordinates (-1.248, -0.914), scale (1:1331083), and other settings.



127.0.0.1:52523/browser/

pgAdmin 4

File Object Tools Help

Browser Dashboard Properties SQL Statistics Dependencies Dependents hamk_db/hamk_user@hamk_db *

public

- Collations
- Domains
- FTS Configurations
- FTS Dictionaries
- FTS Parsers
- FTS Templates
- Foreign Tables
- Functions
- Materialized Views
- Sequences
- Tables (3)
 - postialueet_2019
 - Columns (112)
 - Constraints
 - Indexes
 - Rules
 - Triggers
 - postialueet_tilast...
 - spatial_ref_sys
- Trigger Functions
- Types
- Views

Login/Group Roles

Tablespaces

hamk_db/hamk_user@hamk_db

Query Editor Query History Scratch Pad

```
1 SELECT posti_alue, nimi, tr_mtu, te_as_valj
2 FROM postialueet_2019
3 WHERE te_as_valj < 39 AND tr_mtu < 20000
4 ORDER BY tr_mtu desc;
```

Data Output Explain Messages Notifications

ERROR: permission denied for relation postialueet_2019
SQL state: 42501



HARJOITUS 3: TIETOJEN HAKEMINEN TIETOKANNASTA

Kokeillaan SQL:n kirjoittamista selaimessa:

https://www.w3schools.com/sql/trysql.asp?filename=trysql_asc

Harjoitellaan datan hakemista tietokannasta opettajan ohjeiden mukaan pgAdminilla.



KÄYTÄNNÖN ESIMERKIT

- Keskustelua: mitä päivän sisällöstä jäi päällimmäisenä mieleen?
Mikä jäi epäselväksi?
- Kysymyksiä päivän sisällöstä?



KOTITEHTÄVÄT

- Vastaa kysymyksiin verkkokurssilla
- Vastaa palautteeseen
- Mieti miten voisit hyödyntää avointa dataa ja eri tietolähteitä omassa liiketoiminnassa / organisaatioissasi



SEURAAVA KOULUTUSPÄIVÄ JA AIHEET

3. päivä: 4.6.2019

Visualisoidaan tietoa

Datan käsittelyn perusteet

- Tiedon laatu, ominaisuustiedot, käsittely

- Tiedon luokittelun periaatteet

Tiedon visualisoinnin mahdollisuudet

Esimerkkejä visualisointimahdollisuuksista

- Liikkuva data (aika ja paikka)

- Työkalut visualisointiin