An introduction to ODISSEI

Tom Emery | 5 November 2019 | CESSDA Widening Skopje 2019

Who are we?

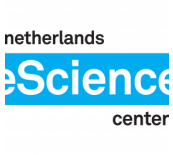
ODISSEI has 32 members that together contribute €1.2 million per year. These include universities, research institutes, ministries and government research agencies. The number of members is continuously growing and diversifying.

Who are we?


Data Archiving and Networked Services
DANS


LISS
PANEL


SURF SARA

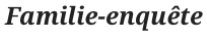

netherlands
eScience
center


trails


European Values Study


European
Social
Survey


VM VEILIGHEIDSMONITOR


Familie-enquête

Nationaal Kiezersonderzoek


CONAMORE


GGP


RADAR


NRO
Nationaal Cohortonderzoek Onderwijs


ISSP


DNB Household Survey


SHARE


SOCON-enquête
Sociaal-culturele ontwikkelingen in Nederland


NELLS


CILS4EU
Children of Immigrants
Longitudinal Survey
in Four European
Countries


hbsc


ILASA
Longitudinal
Aging
Study
Amsterdam

CBS Microdata

- Individual Population Register
- Organizational Registers (Companies, Charities etc)
- Tax Records

- High quality Persistent Identifiers everywhere
- Pseudo Anonymized data is available for research

SURF SARA

What do we do?

Microdata Services
Secure access to CBS data for around 700 researchers from a broad range of scientific disciplines

LISS Panel
A national labortaory for high quality methodogical and scientific innovations.

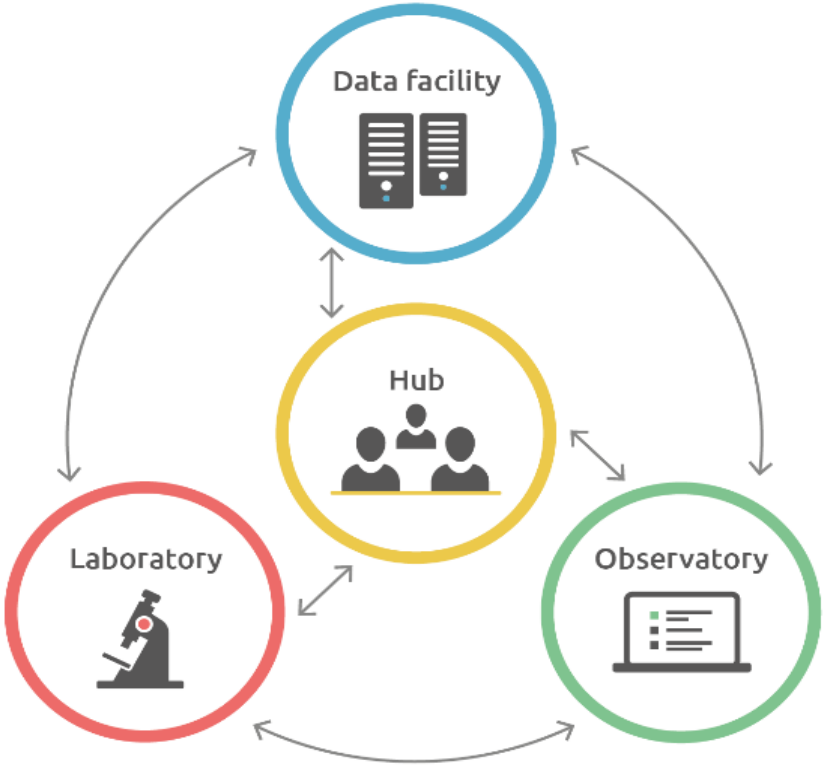
Financing Fieldwork
Maintaining existing panels such as SHARE, ESS, EVS and GGP. Promoting open access data collections in the social sciences with a good price-quality ratio.

HPC Access
Development of a secure HPC environment for social scientists for the analysis of survey and register data.

Standards
Definition and implementation of standards within the social sciences to increase interoperability, monitoring change and improving data quality.

Communications
Building bridges with other disciplines, representing the interests of social scientists and participating in international inittatives.

ODISSEI Roadmap



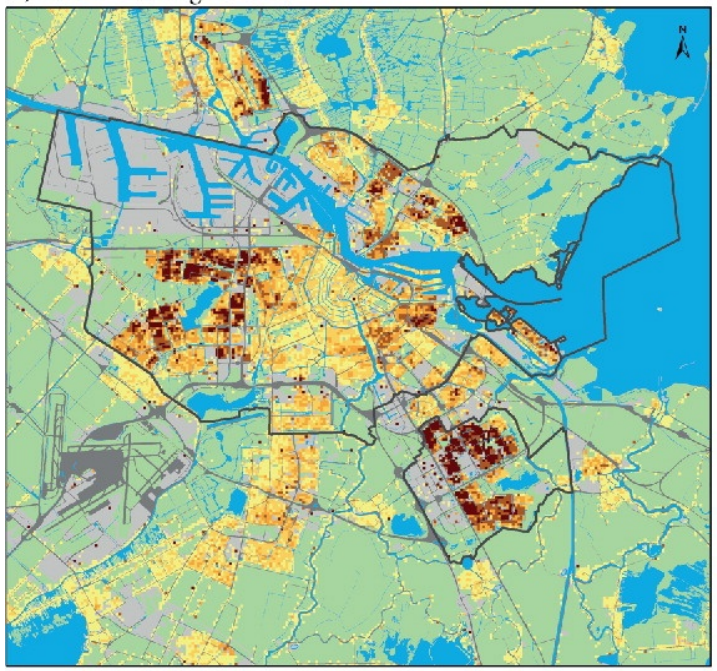
What can we do?

Use Cases in the ODF

Geo-Spatial Analysis

- Precise mapping of specific populations or variables
- For example, % of population with ‘Non-Western’ heritage
- Calculated at small scale 100m x 100m (n = 308,010)
- Calculation repeated for increasing catchment area to measure entropy

A) 100m×100m grid cells



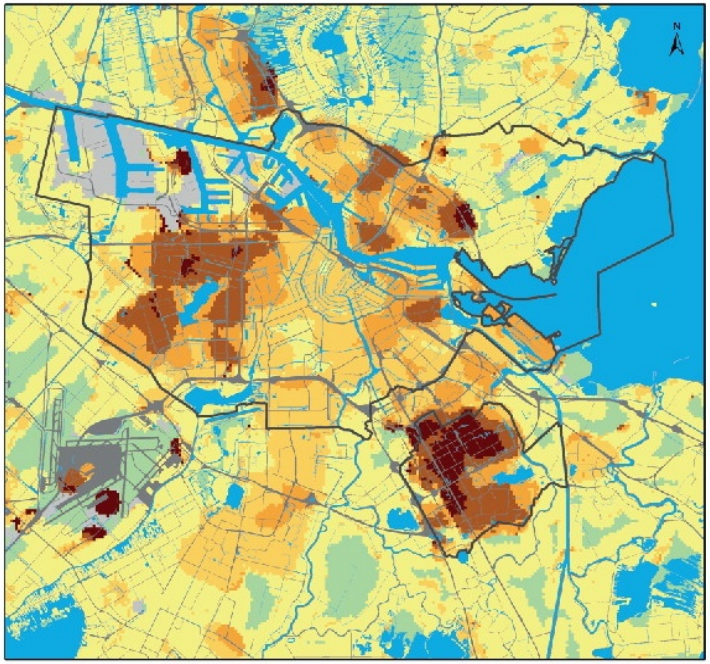
Share of people with a non-Western background (%):

- 0-10
- 11-26
- 27-46
- 47-69
- 70-100

- Municipality of Amsterdam
- Water
- Transportation site
- Built environment
- Green space

Authors' calculations based on data from the Statistics Netherlands (CBS)

B) 500m radius



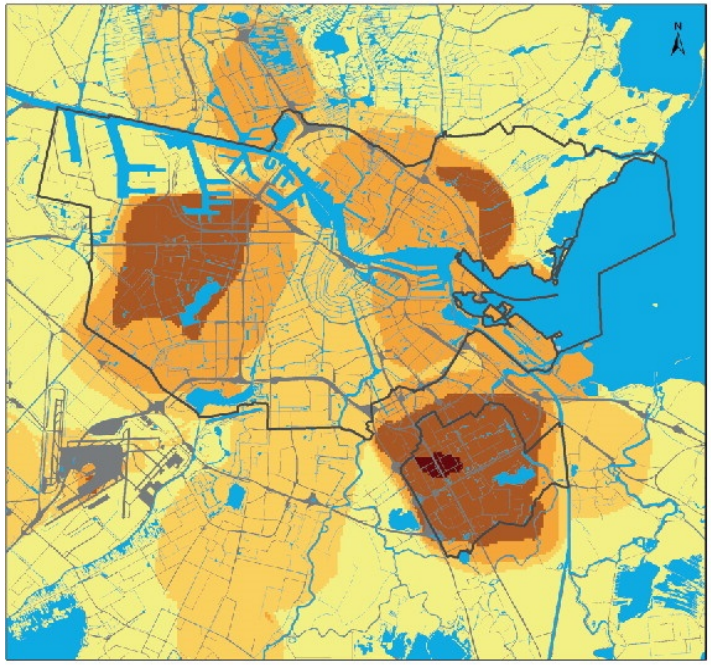
Share of people with a non-Western background (%):

- 0-10
- 11-26
- 27-46
- 47-69
- 70-100

- Municipality of Amsterdam
- Water
- Transportation site
- Built environment
- Green space

Authors' calculations based on data from the Statistics Netherlands (CBS)

C) 2000m radius

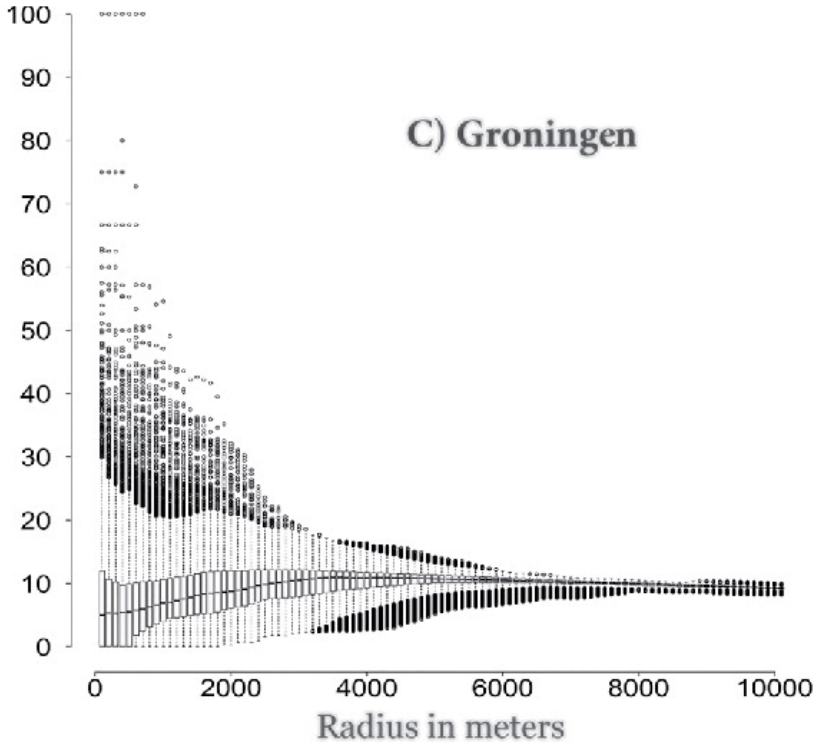


Share of people with a non-Western background (%):

- 0-10
- 11-26
- 27-46
- 47-69
- 70-100

- Municipality of Amsterdam
- Water
- Transportation site
- Built environment
- Green space

Authors' calculations based on data from the Statistics Netherlands (CBS)


C) Groningen
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
Radius in meters

Geo-Spatial Analyis

- This took four months to run at CBS on one Node
- ODF allows us to distribute the job
- This allows analysis to expand to temporal analysis multiple indicators
- Resulting data can also be integrated into any other ODF analysis

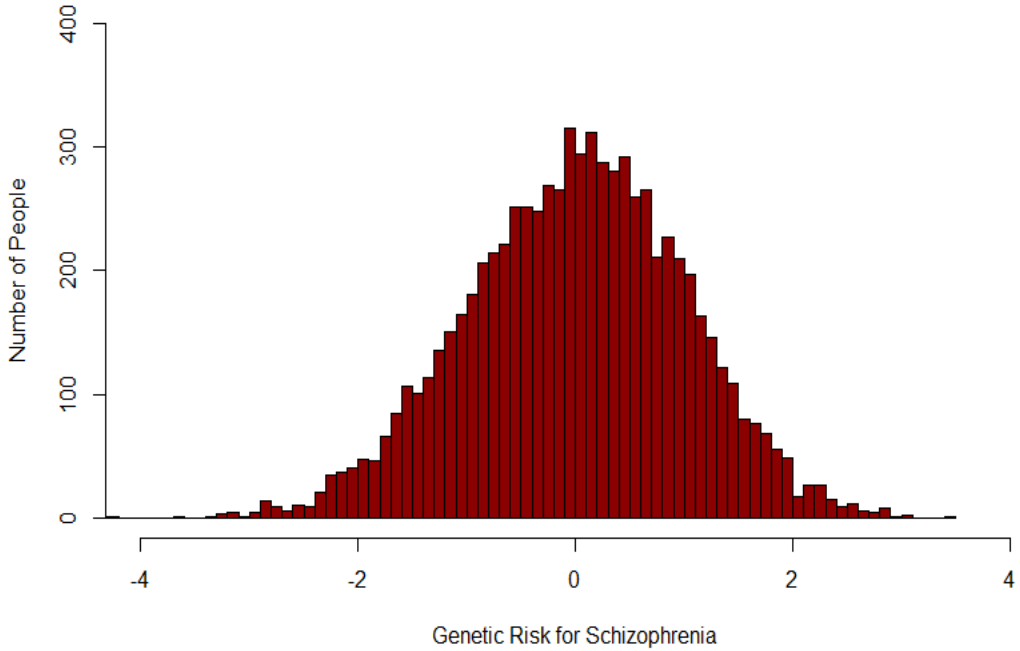
Petrović, Ana, Maarten van Ham, and David Manley. "Multiscale Measures of Population: Within- and between-City Variation in Exposure to the Sociospatial Context." *Annals of the American Association of Geographers* (2018): 1-18.

Genetic Analysis

- National Twin Registry (11,000 Genetic Samples)
- Linking Genetic Data to Registers (with Consent!)
- Data is big, computation is demanding (GWAS)
- One persons genetic data is bigger than the register

- Too big for CBS

Genetic Analysis



Genetic Analysis

Median genetic risk score for schizophrenia in Dutch Towns

Genetic Analysis

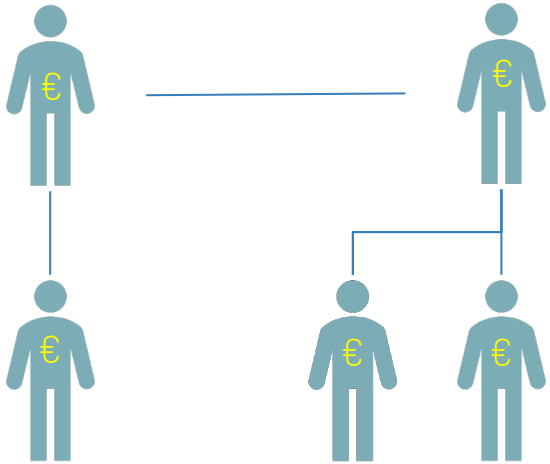
- Genetic Analysis is already done at SurfSara
- The ODF can allow the enrichment of medical data with Administrative data
- Better environment for interdisciplinary research

Colodro-Conde, Lucía et al. 2018. "Association Between Population Density and Genetic Risk for Schizophrenia." JAMA Psychiatry. Retrieved July 5, 2018 (http://archpsyc.jamanetwork.com/article.aspx?doi=10.1001/jamapsychiatry.2018.1581).

Social Network Analysis

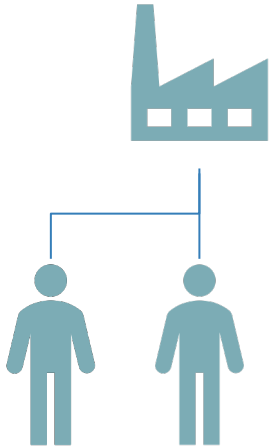
- CBS data used for demographic research to match partners or parents and children etc.
- But… the main constraint is computational
- Small scale sub sample analysis: 4,000 links per person
- Scaling up would be 72 billion links for the whole Dutch population

Social Network Analysis



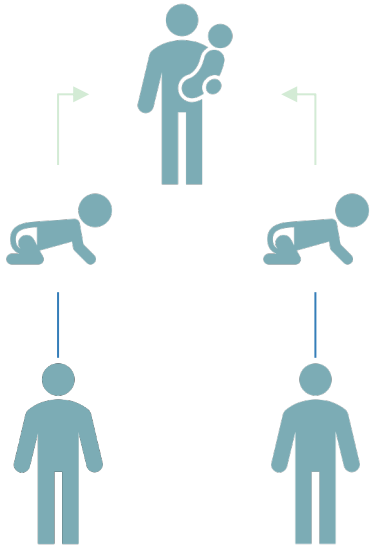
Family Networks

Social Network Analysis



Colleague Networks

Social Network Analysis



School Networks

Social Network Analysis

- Many, many possibilities
- Network Linkages made can be reused in the ODF
- E.g. calculating entropy of specific characteristics across social network and geospatial dimensions simultaneously

Thankyou for your attention

https://odissei-data.nl