ODISSEI

Open Data Infrastructure for Social Science and Economic Innovations

The ODISSEI Secure Supercomputer

Annette Langedijk, Lucas van der Meer ▪ ICTeSSH ▪ 30 June 2020

ODISSEI

Open Data Infrastructure for Social Science and Economic Innovations

Who are we?

Annette Langedijk

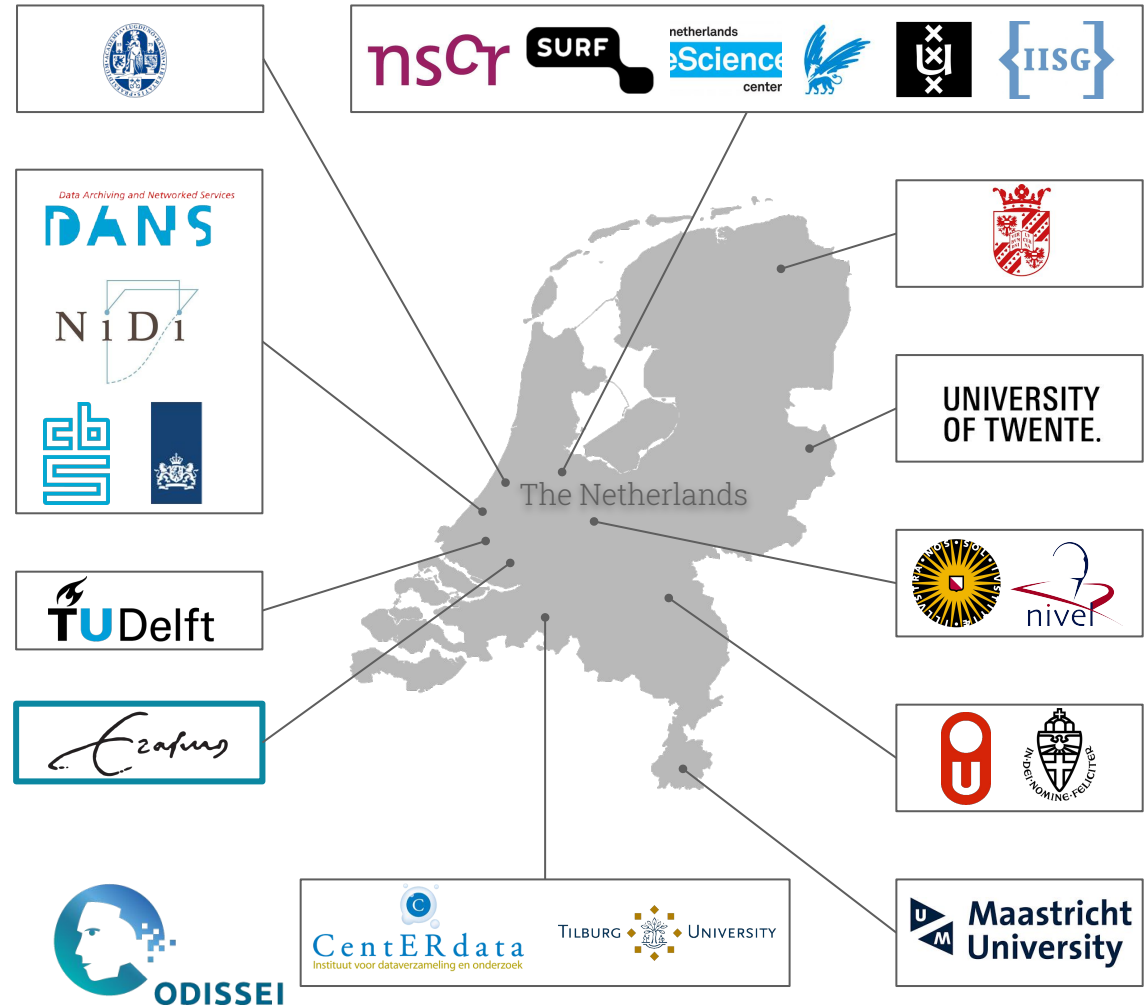
Community manager SSH SURF

Management Board ODISSEI

Lucas van der Meer

Operational manager ODISSEI

ODISSEI

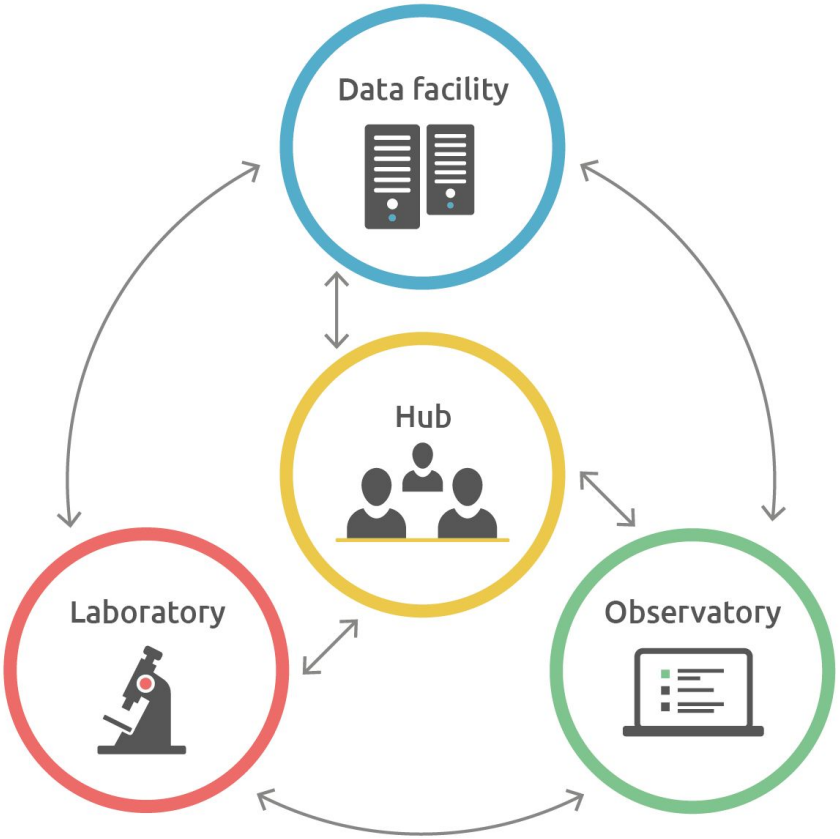


A unique non-profit collaborative consortium

ODISSEI creates a federated data infrastructure for the social and economic sciences in the Netherlands,

on behalf of 40 member organisations.

The ODISSEI research infrastructure



ODISSEI

From traditional 2D survey data...

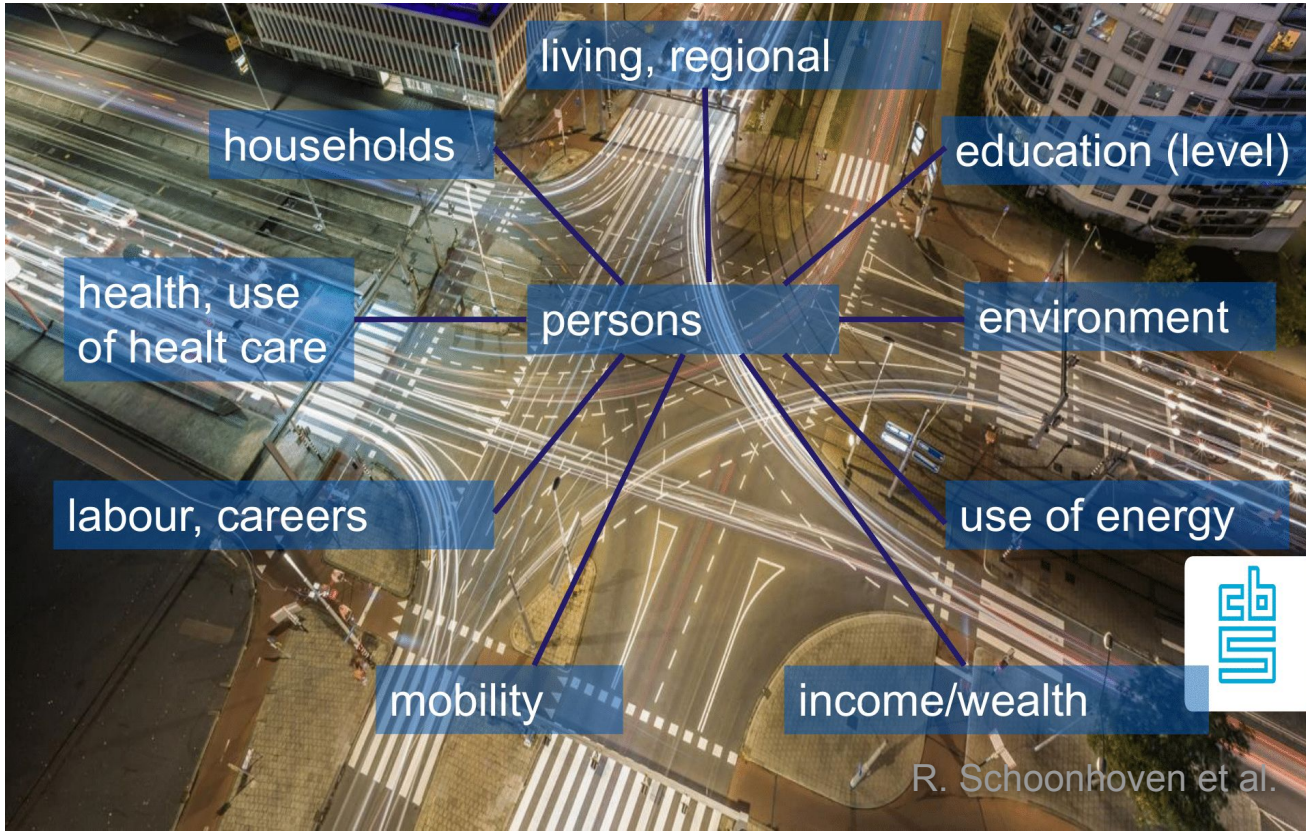
Person#	Sex	Occupation	...
454a87	M	Unemployed	
986c77	F	Plumber	
...			

ODISSEI

From traditional 2D survey data...

Person#	Sex	Occupation	...
454a87	M	Unemployed	
986c77	F	Plumber	
...			

...to rich multi-dimensional, linkable data



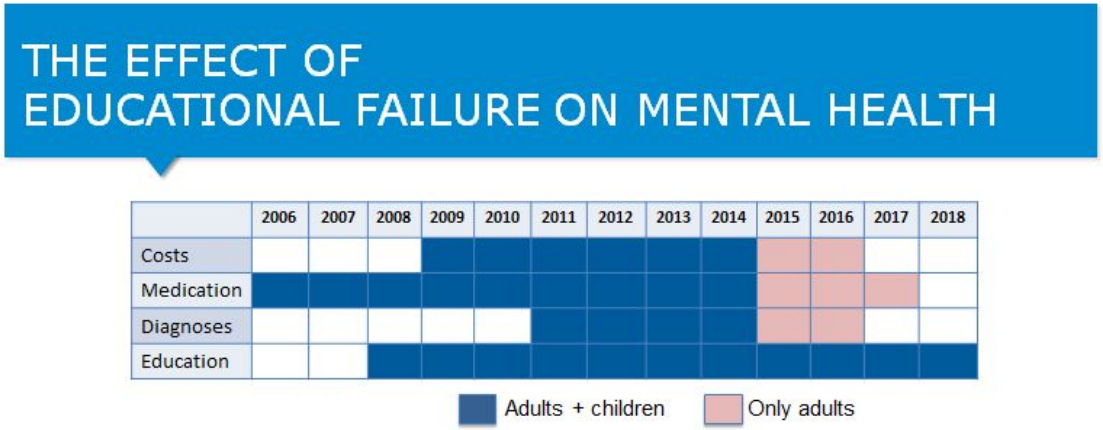
ODISSEI

Microdata from Statistics Netherlands (CBS)

Microdata:

linkable data on individuals, companies, and addresses

Example Microdata research

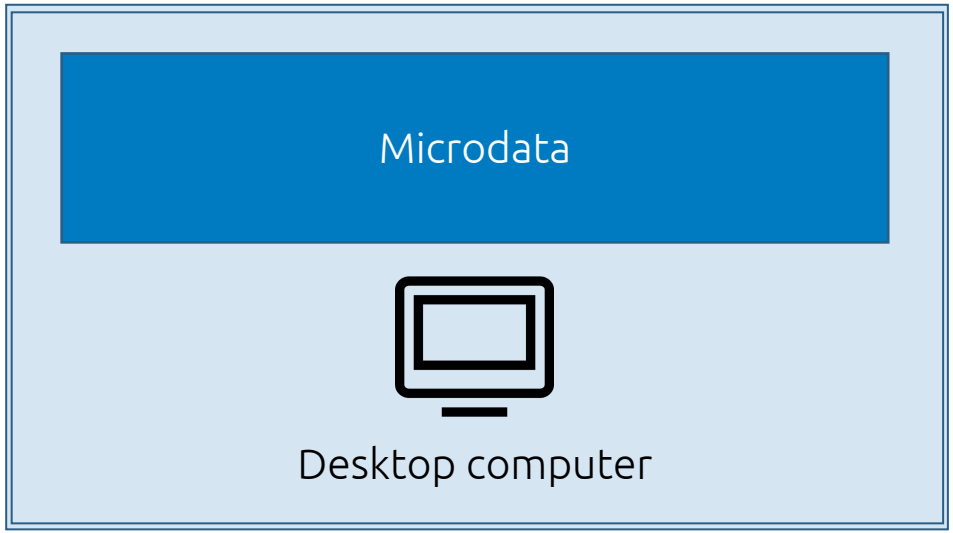


VU VRIJE UNIVERSITEIT AMSTERDAM

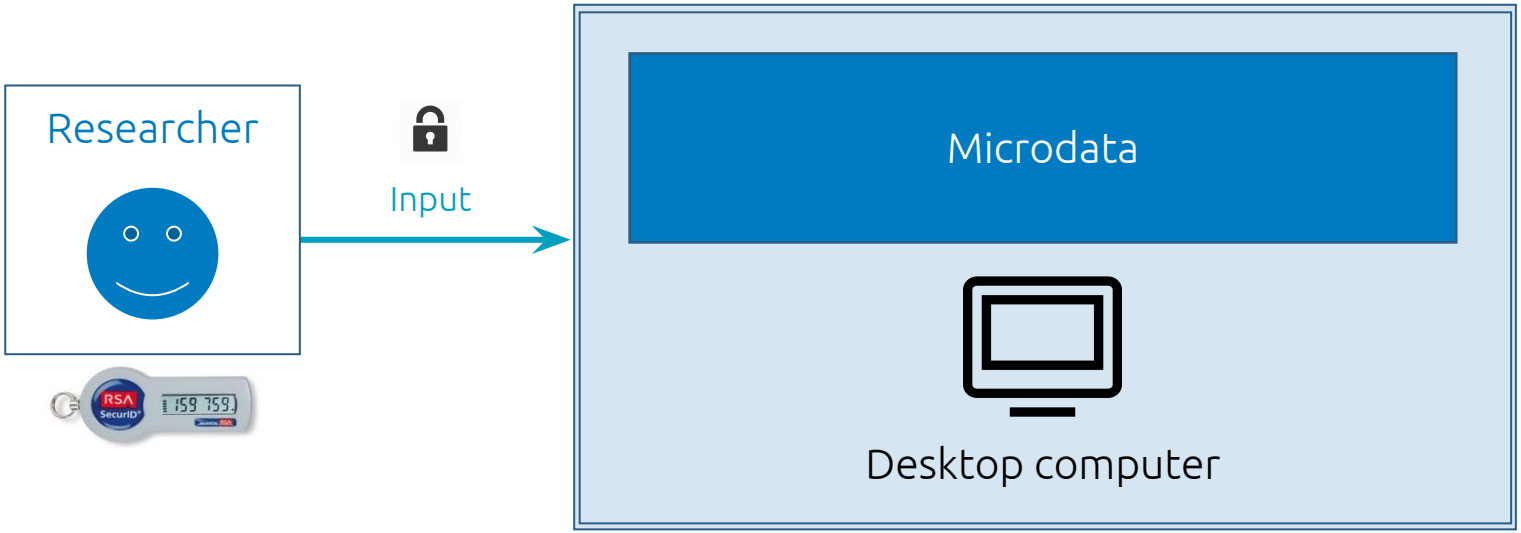
P. Demange et al.

ODISSEI

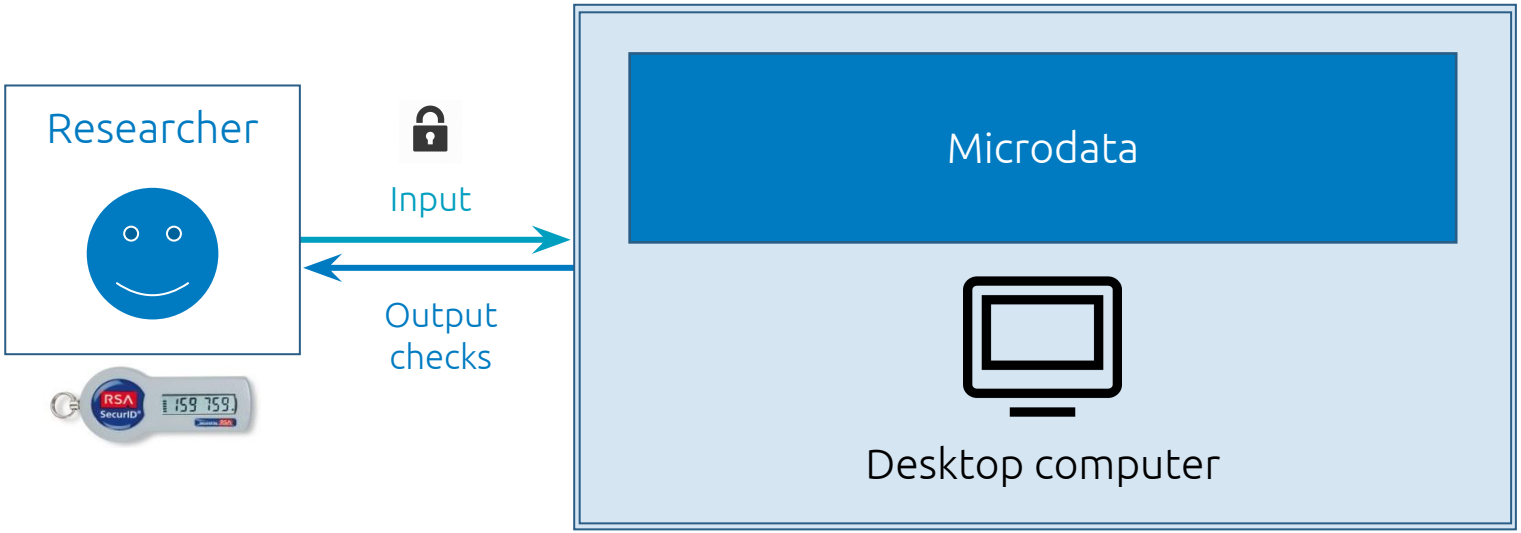
Statistics NL Remote Access environment



Statistics NL Remote Access environment



Statistics NL Remote Access environment



Challenge: billions of data points

585,000 populated cells
× 101 scales × 2 variables
× 15 years = 1.8 billion
data points

⟶ 4 months continuous
calculations

ODISSEI Secure Supercomputer

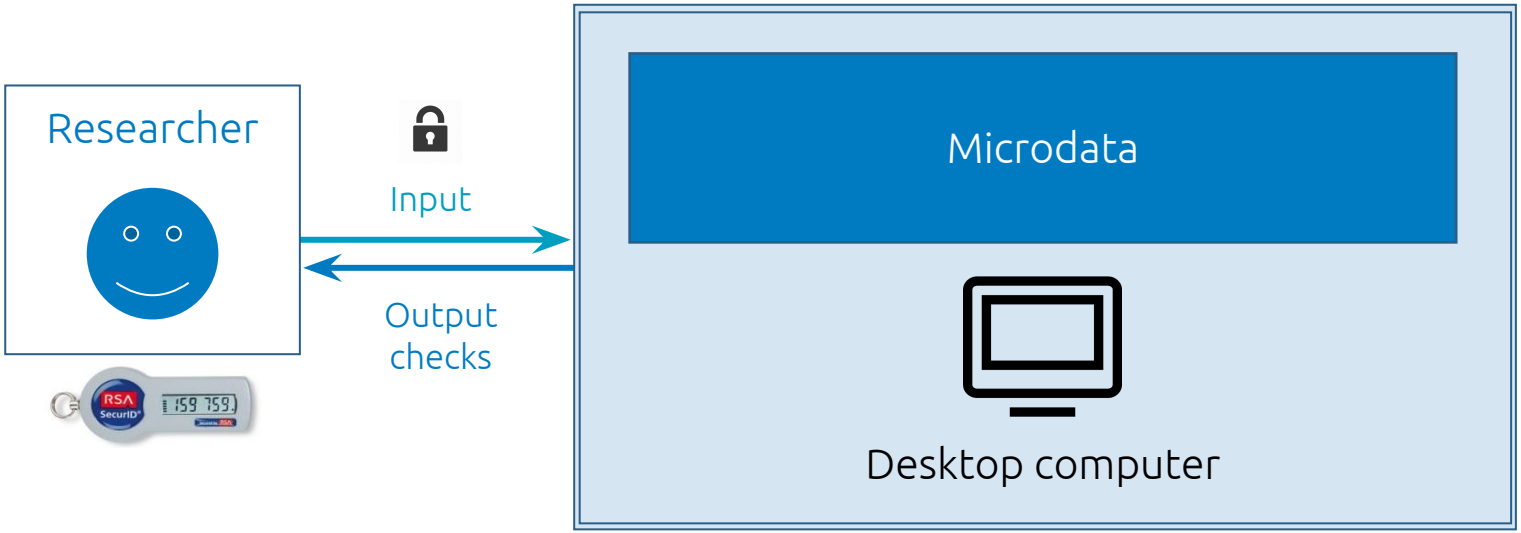
Open Data Infrastructure for Social Science and Economic Innovations

Cartesius, the Dutch National supercomputer

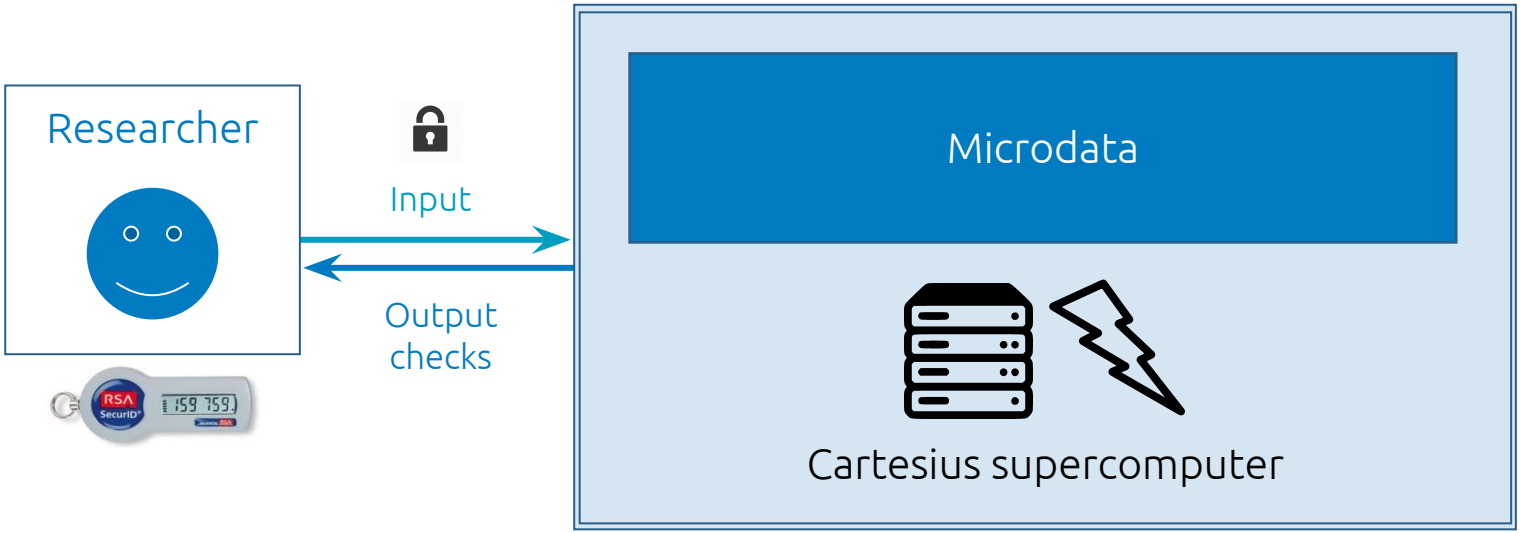
The collaborative organisation for ICT in Dutch education and research

ODISSEI

Statistics NL Remote Access environment

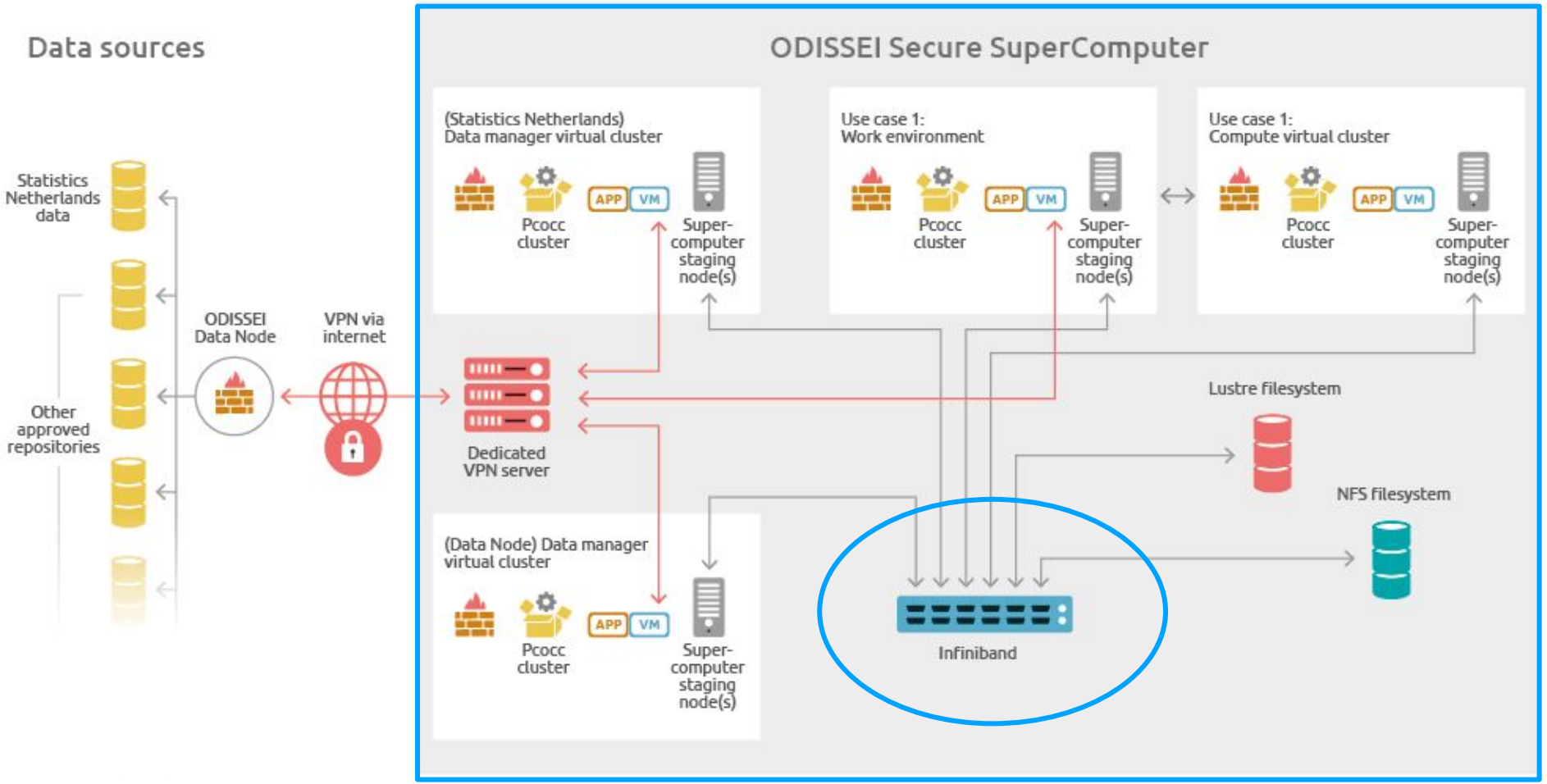


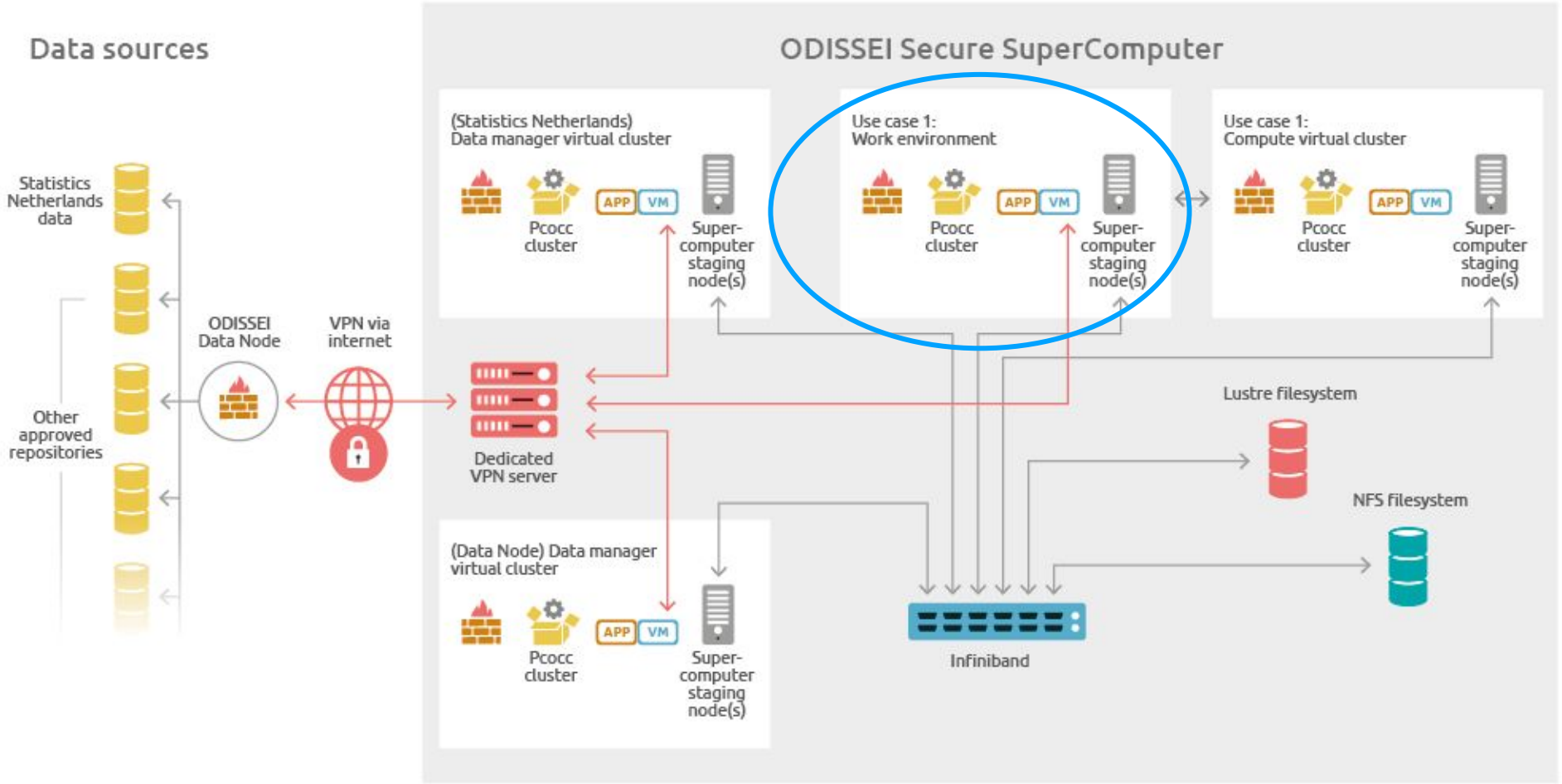
The ODISSEI Secure Supercomputer (OSSC)

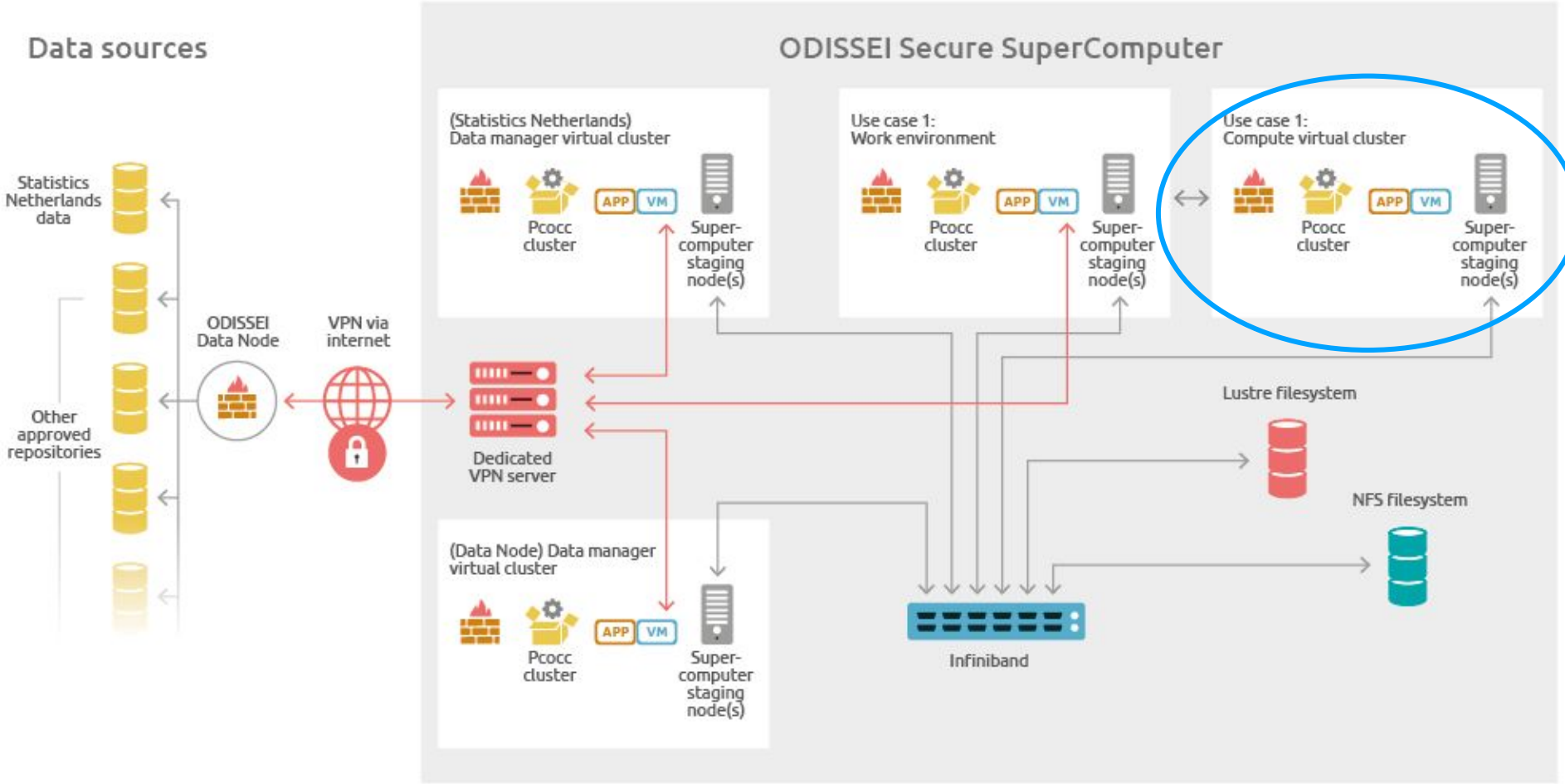


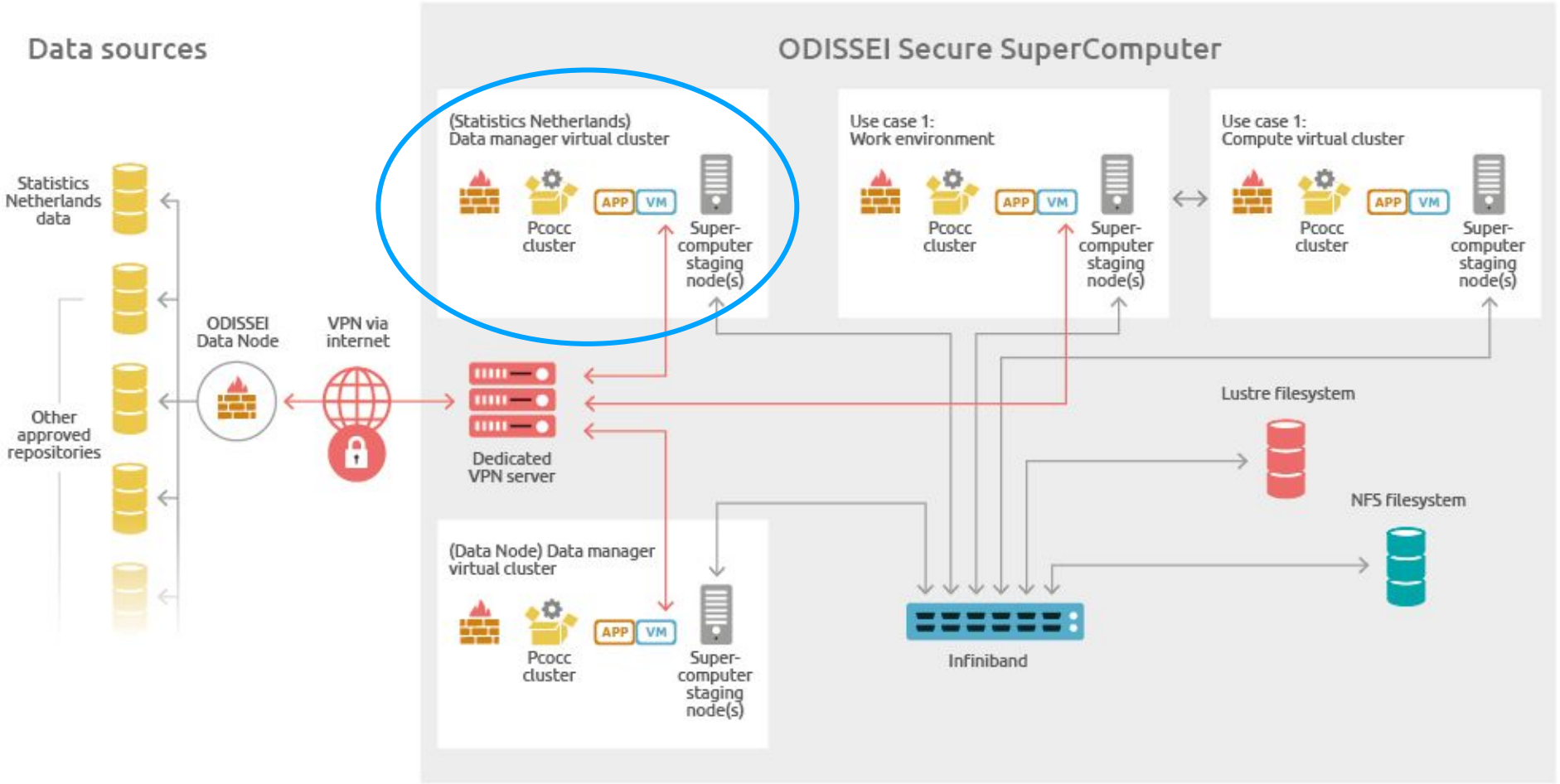
ODISSEI

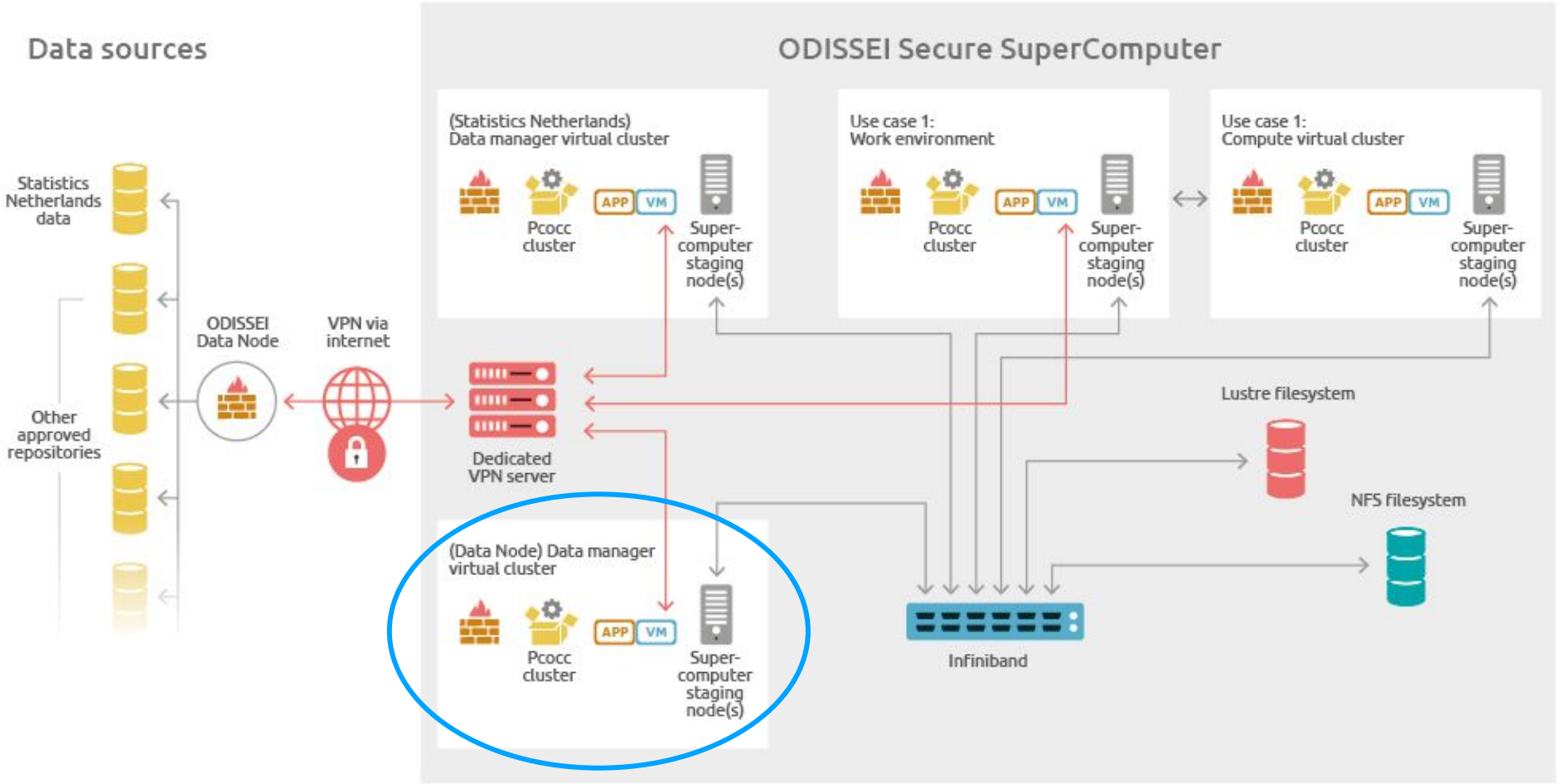
SURF

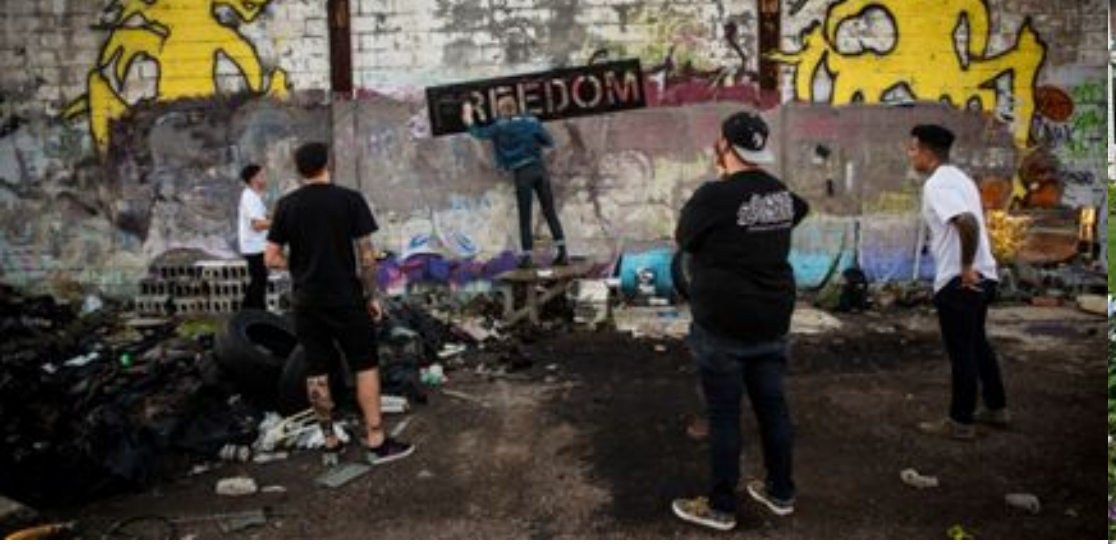

Data sources
ODISSEI Secure SuperComputer
Statistics Netherlands data
Other approved repositories
ODISSEI Data Node
VPN via internet
Dedicated VPN server
(Statistics Netherlands) Data manager virtual cluster
Pcocc cluster
APP
VM
Super-computer staging node(s)
Use case 1: Work environment
Use case 1: Compute virtual cluster
(Data Node) Data manager virtual cluster
Lustre filesystem
NFS filesystem
Infiniband


Data sources
ODISSEI Secure SuperComputer
Statistics Netherlands data
Other approved repositories
ODISSEI Data Node
VPN via internet
Dedicated VPN server
(Statistics Netherlands) Data manager virtual cluster
Pcocc cluster
APP
VM
Super-computer staging node(s)
Use case 1: Work environment
Use case 1: Compute virtual cluster
(Data Node) Data manager virtual cluster
Lustre filesystem
NFS filesystem
Infiniband


Data sources
ODISSEI Secure SuperComputer
Statistics Netherlands data
Other approved repositories
ODISSEI Data Node
VPN via internet
Dedicated VPN server
(Statistics Netherlands) Data manager virtual cluster
Pcocc cluster
APP
VM
Super-computer staging node(s)
Use case 1: Work environment
Use case 1: Compute virtual cluster
(Data Node) Data manager virtual cluster
Lustre filesystem
NFS filesystem
Infiniband

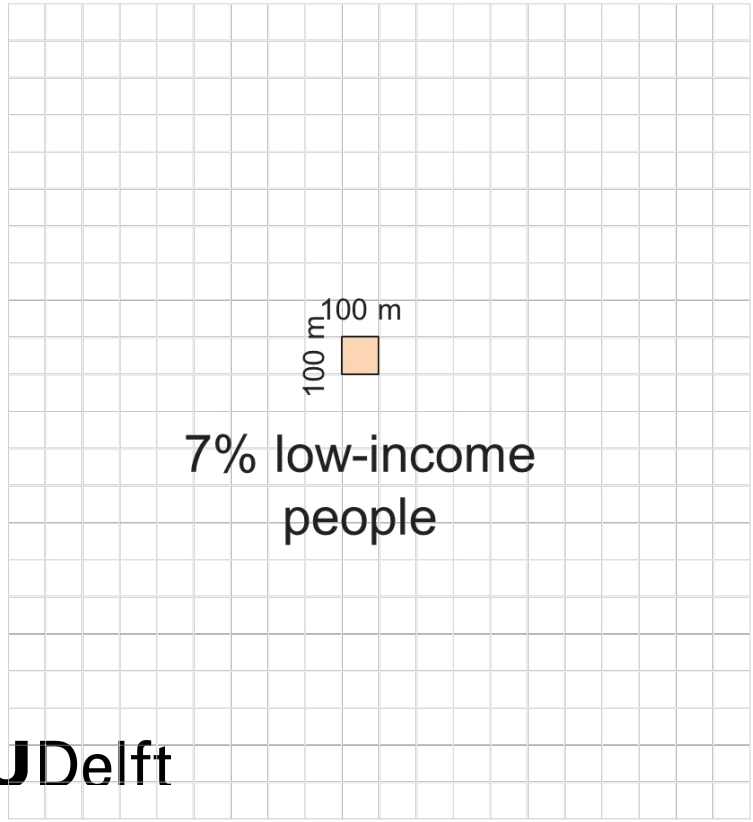

Data sources
ODISSEI Secure SuperComputer
Statistics Netherlands data
Other approved repositories
ODISSEI Data Node
VPN via internet
Dedicated VPN server
(Statistics Netherlands) Data manager virtual cluster
Pcocc cluster
APP
VM
Super-computer staging node(s)
Use case 1: Work environment
Pcocc cluster
APP
VM
Super-computer staging node(s)
Use case 1: Compute virtual cluster
Pcocc cluster
APP
VM
Super-computer staging node(s)
Lustre filesystem
NFS filesystem
(Data Node) Data manager virtual cluster
Pcocc cluster
APP
VM
Super-computer staging node(s)
Infiniband

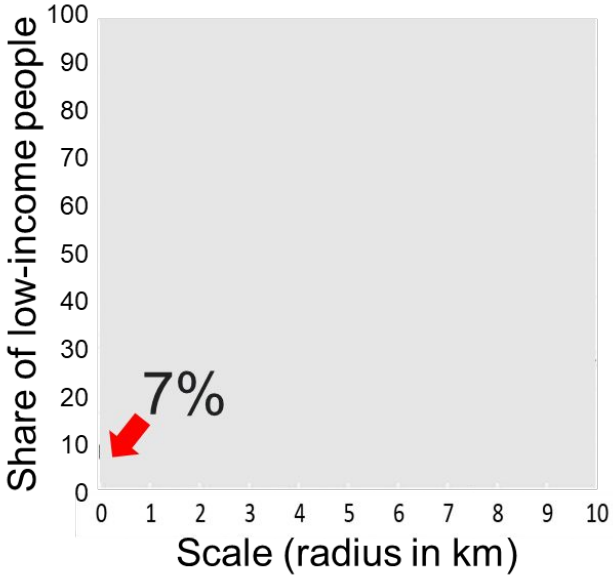

Data sources
ODISSEI Secure SuperComputer
Statistics Netherlands data
Other approved repositories
ODISSEI Data Node
VPN via internet
Dedicated VPN server
(Statistics Netherlands) Data manager virtual cluster
Pcocc cluster
APP
VM
Super-computer staging node(s)
Use case 1: Work environment
Use case 1: Compute virtual cluster
(Data Node) Data manager virtual cluster
Lustre filesystem
NFS filesystem
Infiniband



Use Case:
Contextual effects at multiple spatial scales for the full population of the Netherlands

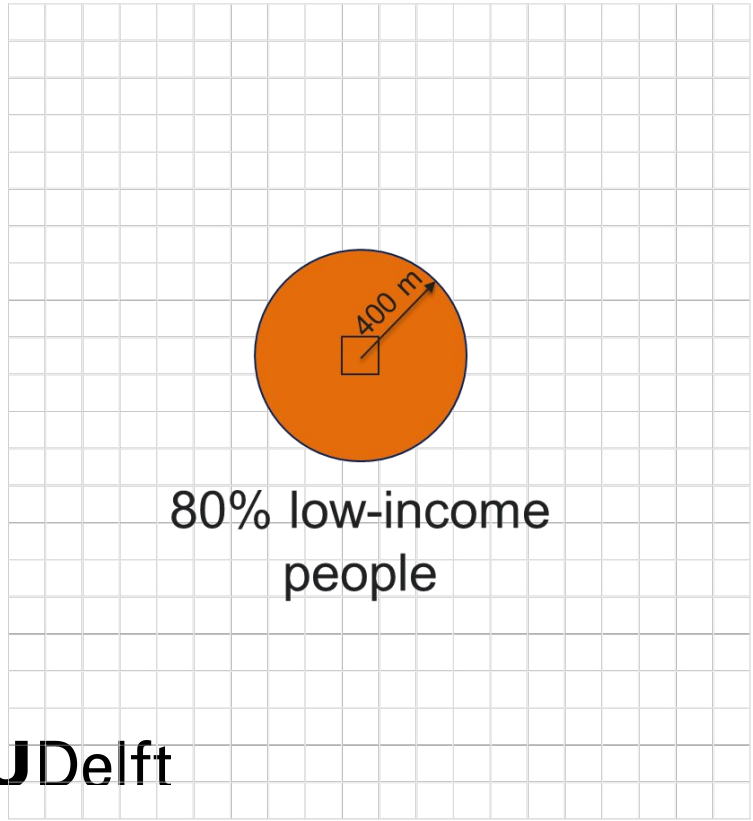
Exposure to spatial context at 101 scales

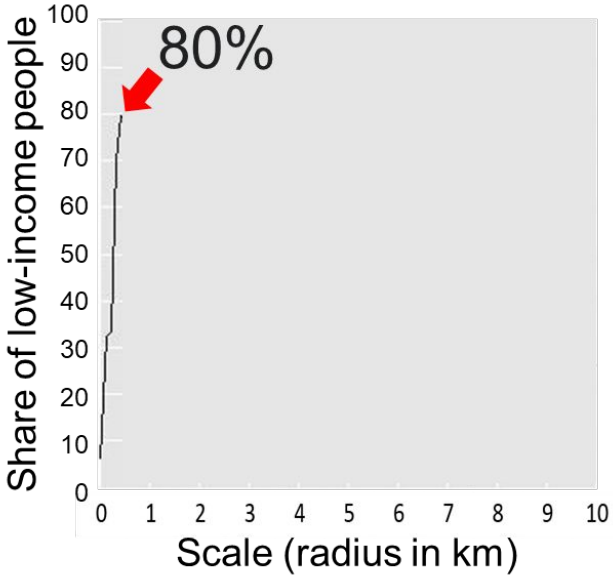




(see Petrović, van Ham, & Manley, 2018)

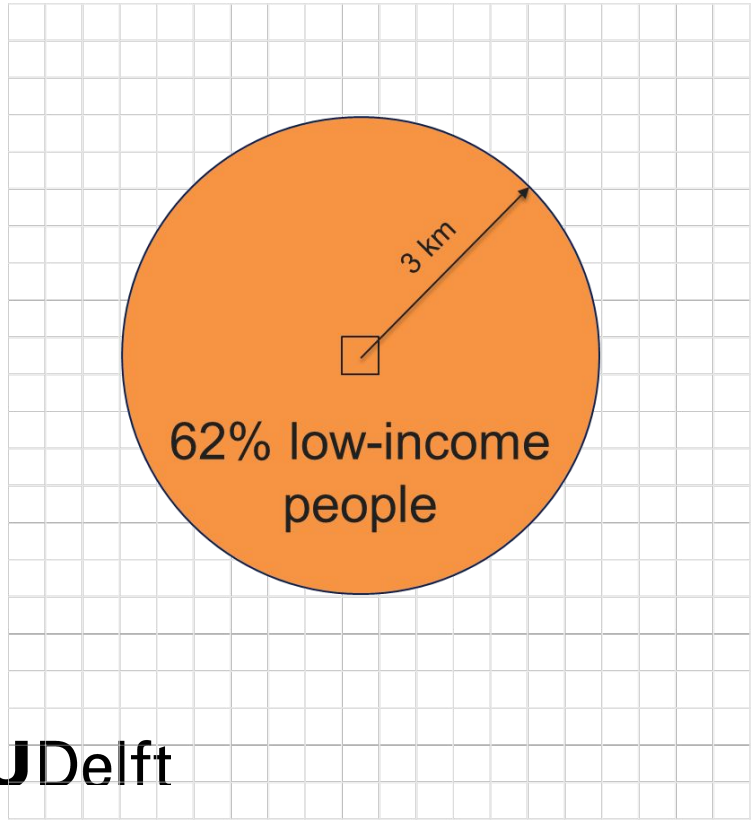
Exposure to spatial context at 101 scales

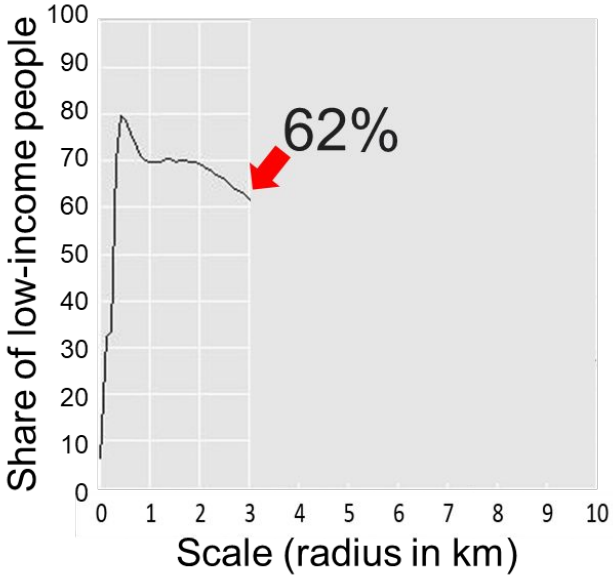




(see Petrović, van Ham, & Manley, 2018)

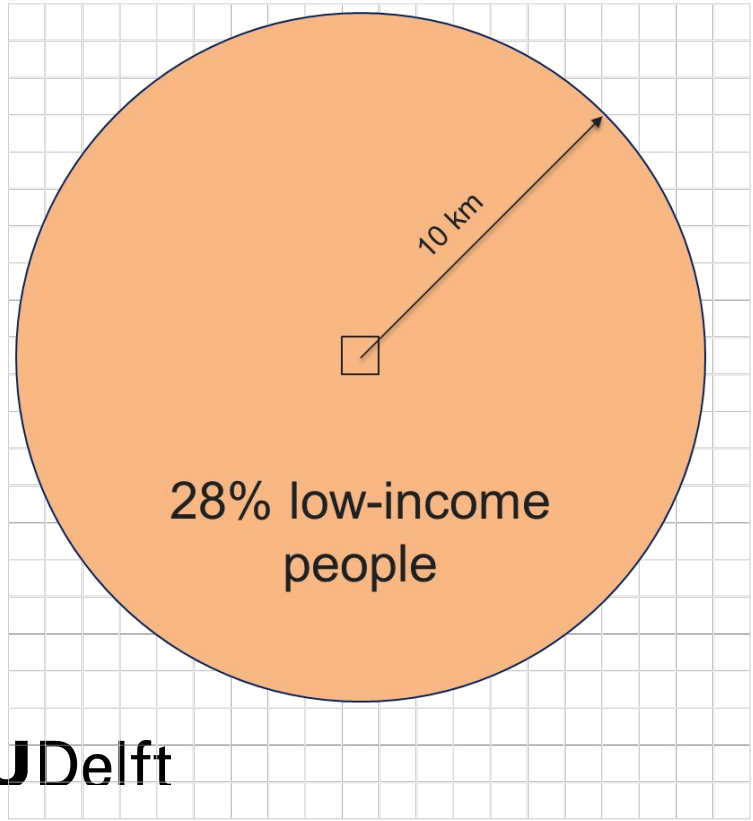
Exposure to spatial context at 101 scales

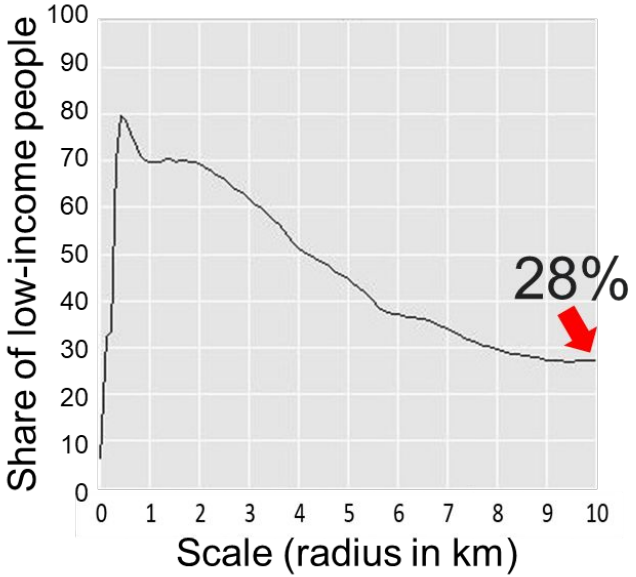




(see Petrović, van Ham, & Manley, 2018)

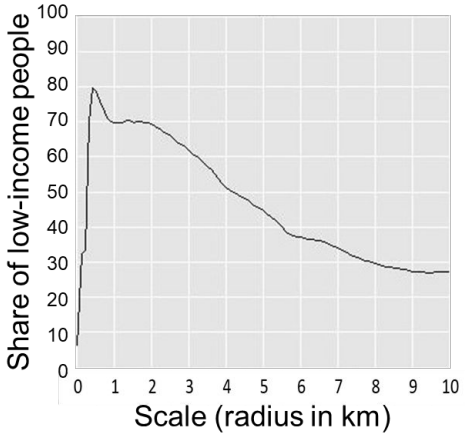
Exposure to spatial context at 101 scales



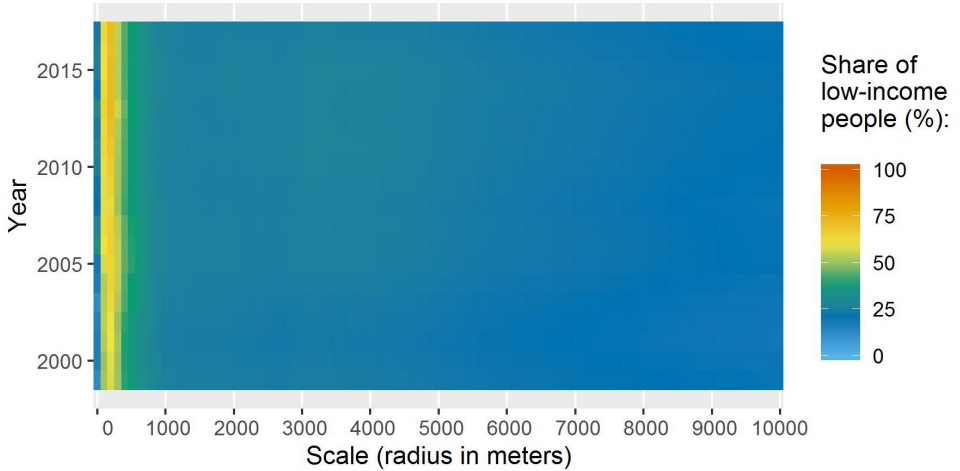


(see Petrović, van Ham, & Manley, 2018)

Share of low-income people in one residential location, measured at 101 scales in 19 years



× 19 =



Challenge: billions of data points

585,000 populated cells
× 101 scales × 2 variables
× 15 years = 1.8 billion
data points

⟶ 4 months continuous
calculations

Reduced computing time on ODISSEI Secure Supercomputer

585,000 populated cells
× 101 scales × 2 variables
× 15 years = 1.8 billion
data points

⟶ 4 months continuous
calculations

585,000 populated cells
× 101 scales × 3 variables
× 19 years = 3.4 billion
data points

⟶ 1 week on 24 nodes

The ODISSEI Secure Supercomputer (OSSC): Unique

Nowhere else is there an infrastructure that allows to analyse register and survey data on a high-performance computer, which is available to the entire community of social scientists.

The ODISSEI Secure Supercomputer (OSSC): Future

The OSSC has been extensively piloted and will be opened up to ODISSEI member organisations in September 2020.

The OSSC design can be adapted to any sensitive data provider, not just Statistics Netherlands. Health insurers, private companies...

Users SSH, builders ICT

- Take time to understand each other
- Plan for early user testing and allow for design adaptations
- Arrange support for researchers (Linux, parallelisation)
- Computational turn in SSH: show, don't tell

ODISSEI

Thank you

info@odissei-data.nl

www.odissei-data.nl/ossc

ODISSEI

Open Data Infrastructure for Social Science and Economic Innovations

The ODISSEI Secure Supercomputer

Annette Langedijk, Lucas van der Meer ▪ ICTeSSH ▪ 30 June 2020

ODISSEI

Open Data Infrastructure for Social Science and Economic Innovations