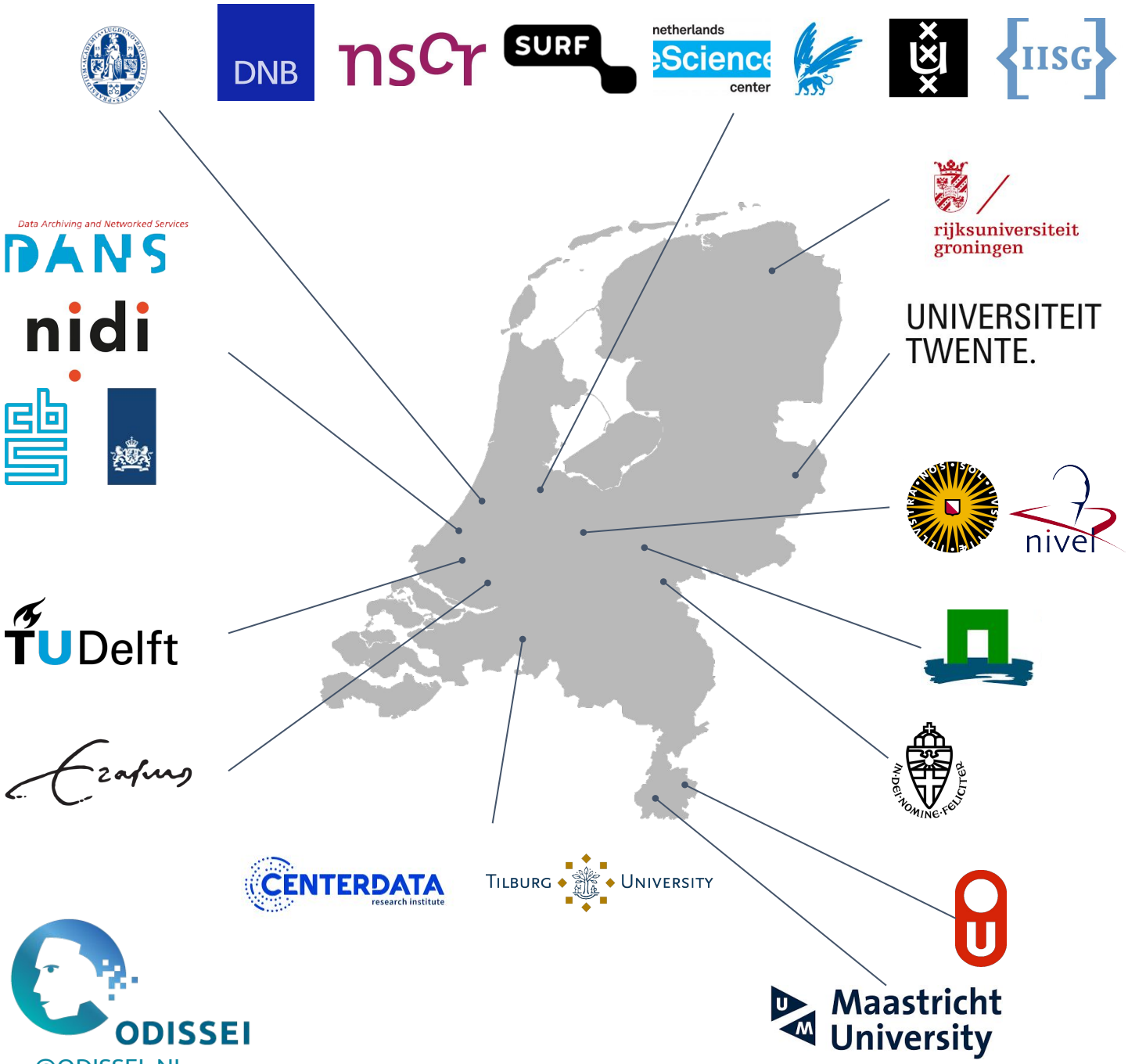
ODISSEI

Open Data Infrastructure for Social Science and Economic Innovations



A national infrastructure for Social Science

ODISSEI creates a federated data infrastructure for the social and economic sciences in the Netherlands, on behalf of more than 40 member organisations.

Research question:
"How do we fight polarisation?"

- Wicked problem
- Needs many researchers – fast
- Needs scalable, interconnected infrastructure
- Machine-readable FAIR

Image: Olaf Kraak / Trouw ©

CBS Microdata

Large source of government registrations:

- Vaccination (GGD)
- Income (Belastingdienst)
- Diploma's (DUO)

Find appropriate data

Via the **ODISSEI Portal**, researchers perform advanced variable-level queries

FAIR metadata

Person# (pseud.)	**Sex**	**Age**	**Marital Status**	**Vaccinated**	**…**
1002020	M	28	Married	0	…
1002021	F	32	Married	1	…
1002022	M	61	Widowed	3	…

CBS Microdata - *n*D

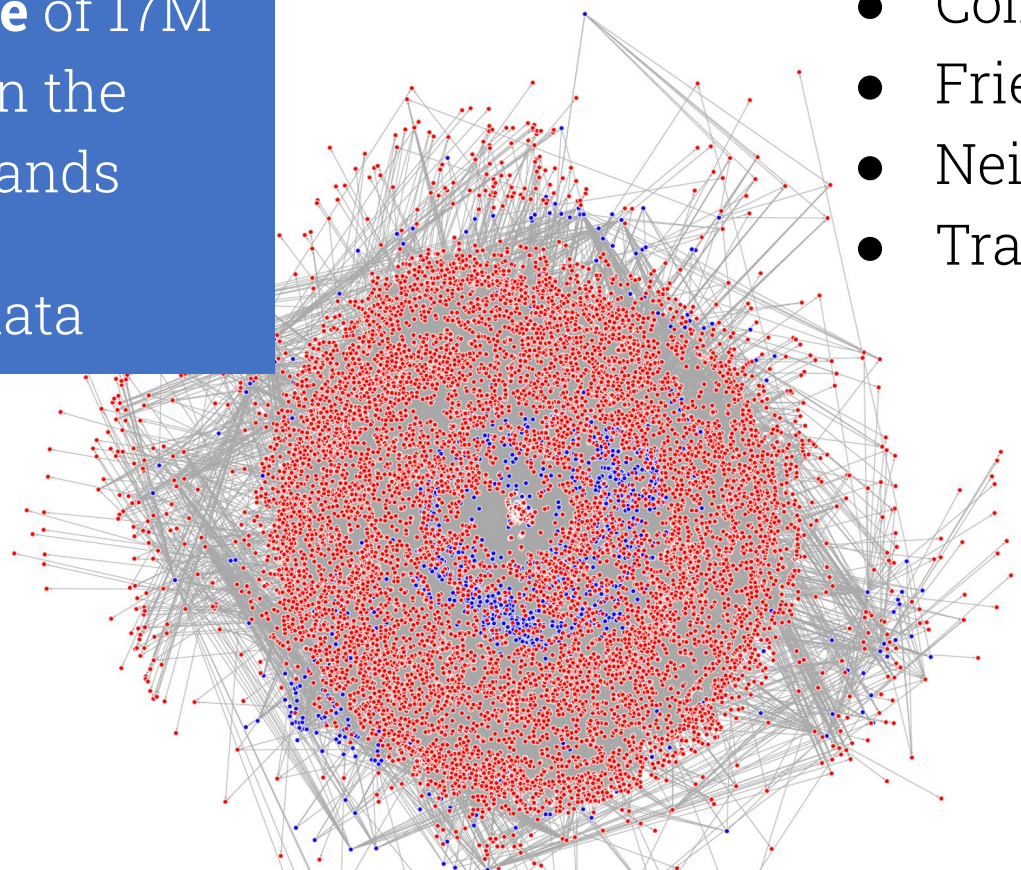
Via ODISSEI, **CBS** constructed a **network file** of 17M people in the Netherlands

FAIR data

Use **interop-erability** between datasets

People and groups are related to each other in lots of ways

- Families
- Colleagues
- Friends
- Neighbourhoods
- Trading



ODISSEI

LISS panel

All CBS pseudonymised persons can be connected to the **LISS panel**

FAIR data

Use **interop-erability** between datasets

No attitudes in CBS → LISS panel:

Periodic survey among ~7,500 people.

E.g. “What is your trust in the government?”

ODISSEI

LISS PANEL

Dutch Parliamentary Election Study (NKO)

Survey about voting behaviour and background information, every parliamentary election.

E.g. “What is your trust in the government?”

Robustness

Use **interop-erability** between questions

Semantic mapping between questions in different surveys in **ODISSEI Portal**

FAIR constructs, measurements, survey questions

ODISSEI

Royal Library (KB), Netherlands Institute for Sound & Vision (B&G)

Use **interop-erability** between datasets

Via the **ODISSEI Portal**, researchers perform semantic queries

FAIR metadata

- Find historical newspaper trends about the topic (not word) “polarisation”
  - E.g. “trust”, “distrust”
- Talk show video performances:
  - how many experts vs. regular people?
  - how many man vs. women?

ODISSEI

Automatic **accessibility** procedures

Data access broker in the **ODISSEI Portal**

FAIR licences & FAIR researchers

Secure ANalysis Environment (SANE)

KB & NISV have copyrighted material – can only be analysed under specific conditions

Metadata about license + researcher

Harmonisation of access conditions

ODISSEI

Machine-readable & understandable code

Reusability of code & models

ODISSEI **GitHub, Zenodo, User policy**

FAIR code & models

Machine-readable

Jupyter / Stata script

- Which assumptions were made?
- Which methodology was followed?
- How were the data pre-processed?
- What was the outcome of my regression analysis?
- What was the standard deviation?
- …

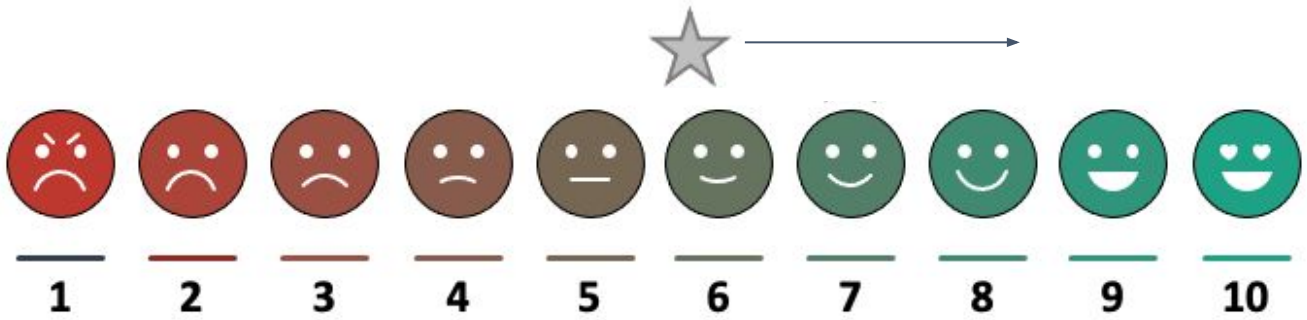
ODISSEI

Research question:

"How do we fight polarisation?"

Image: Olaf Kraak / Trouw ©

There are still many Open challenges for Open Science and FAIR data at ODISSEI



An infrastructure is more than the sum of its parts

ODISSEI

FAIR is not just about data

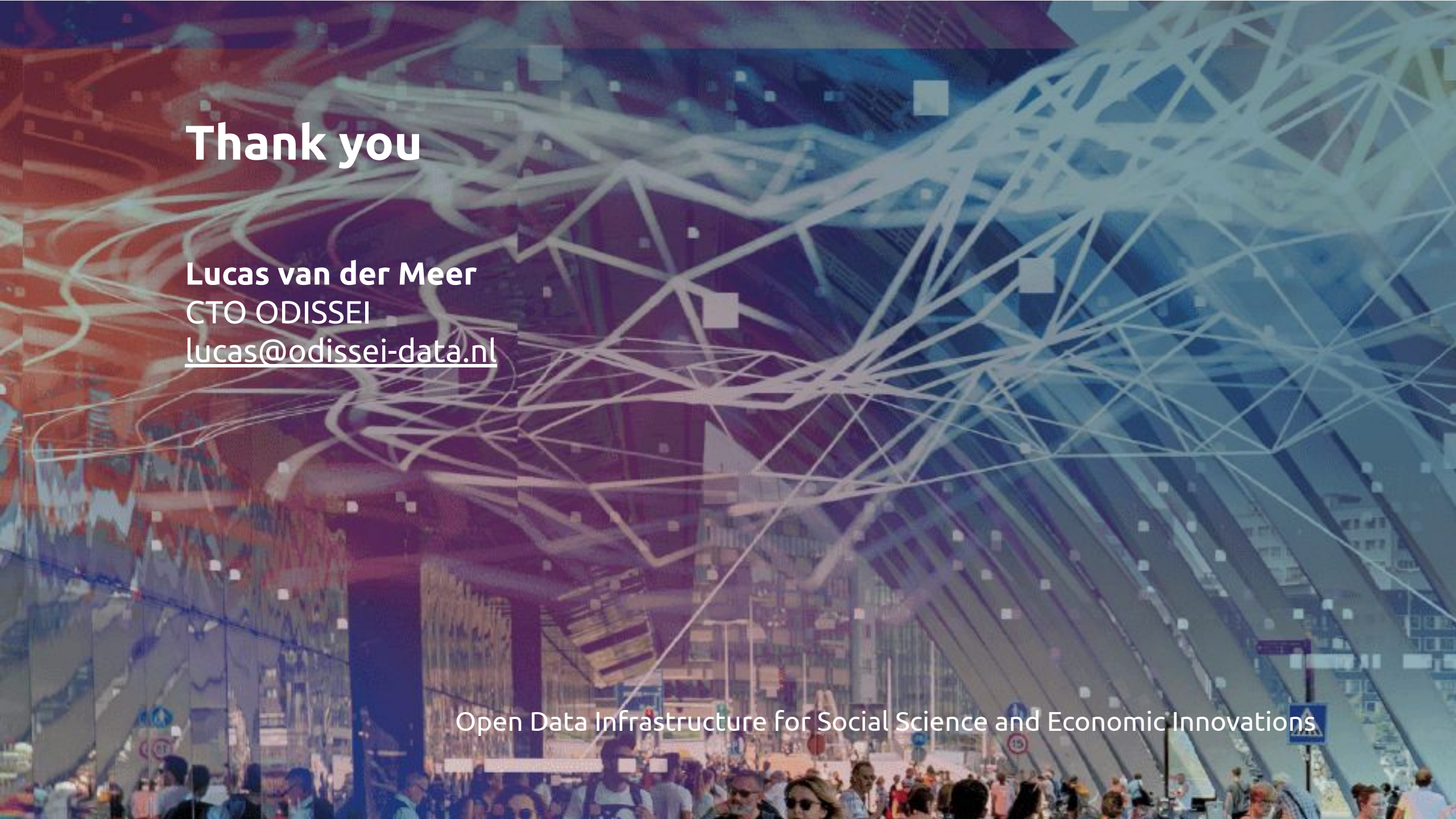
- **Metadata** of datasets (findability)
- **Measurements** of questions in a panel (interoperability)
- **Researcher** properties (accessibility)
- **License** conditions (accessibility)
- **Code** (reusability)

ODISSEI

There are still many Open challenges for Open Science and FAIR data at ODISSEI

- Data sensitivity
  - Access restrictions
  - Public perception (e.g. of using CBS microdata)
- Diversity of data types and stakeholders
  - Lack of standards & interoperability
  - Incentives & motivation
  - Machine-understandability of models

ODISSEI



Thank you

Lucas van der Meer
CTO ODISSEI
lucas@odissei-data.nl

Open Data Infrastructure for Social Science and Economic Innovations