Open Data Reusers in the picture

Dr. Frederika Welle Donker,
f.m.welledonker@tudelft.nl

Open Data Life Cycle Workshop, 30 Oct. 2020

Twinning
Open Data
Operational

Agenda

A **Open Data in the Netherlands**
Policies and portals ▶▶

B **Research overview**
Research questions and methodologies ▶▶

C **Open data users**
Per portal and combined ▶▶

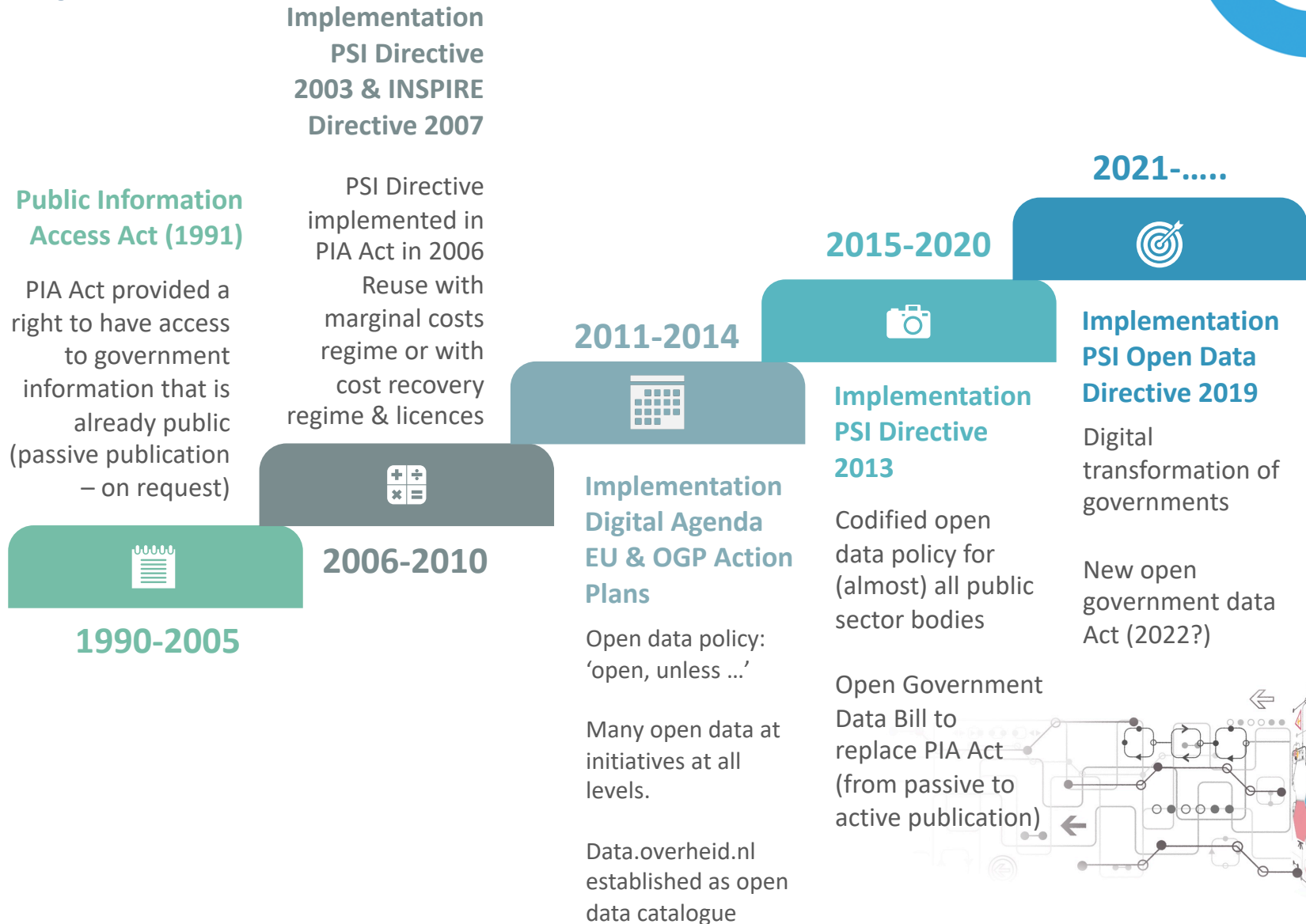**D** **Barriers for OD users**
And how to tackle these barriers ▶▶

E **Conclusions**
Recommendations for future research ▶▶
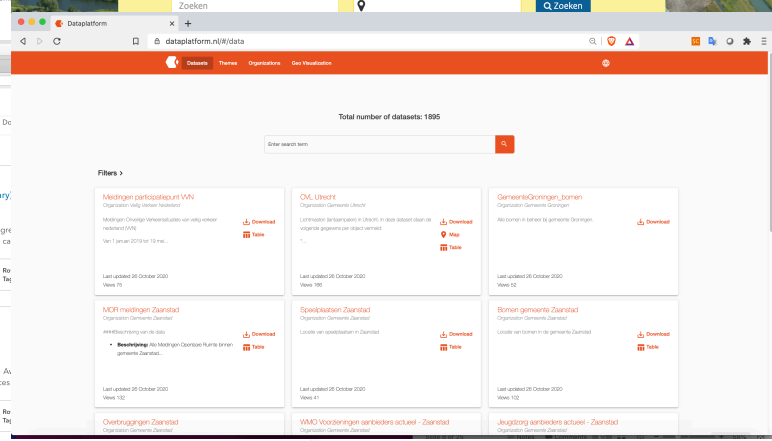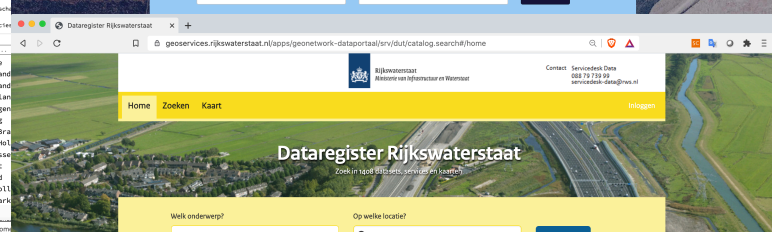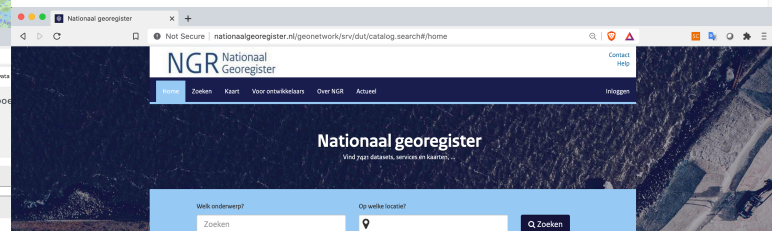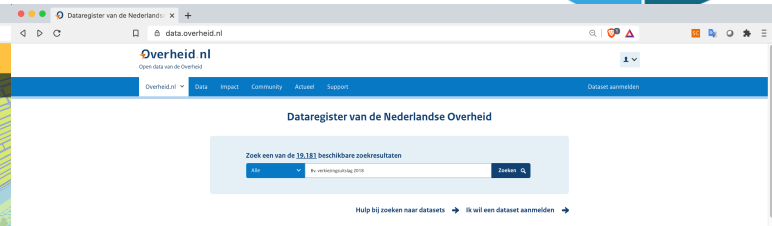
Open Data in the Netherlands

Implementation PSI Directive 2003 & INSPIRE Directive 2007

Public Information Access Act (1991)

PIA Act provided a right to have access to government information that is already public (passive publication – on request)

PSI Directive implemented in PIA Act in 2006 Reuse with marginal costs regime or with cost recovery regime & licences

1990-2005

2006-2010

2011-2014

Implementation Digital Agenda EU & OGP Action Plans

Open data policy: 'open, unless …'

Many open data at initiatives at all levels.

Data.overheid.nl established as open data catalogue

2015-2020

Implementation PSI Directive 2013

Codified open data policy for (almost) all public sector bodies

Open Government Data Bill to replace PIA Act (from passive to active publication)

2021-…..

Implementation PSI Open Data Directive 2019

Digital transformation of governments

New open government data Act (2022?)

Twinning
Open Data
Operational

3

Many open data portals ….
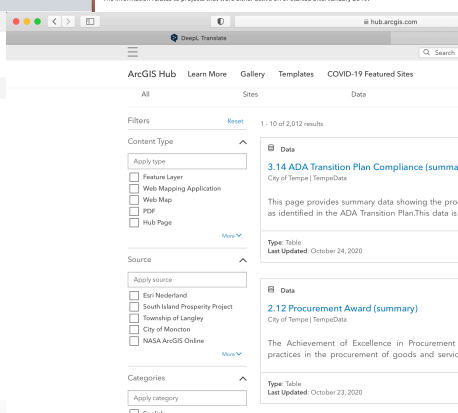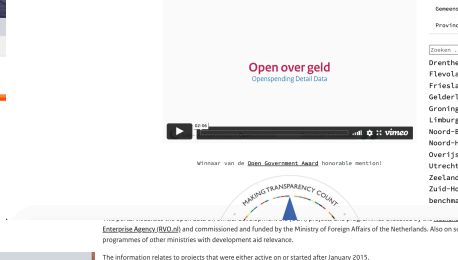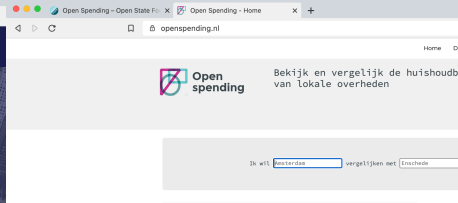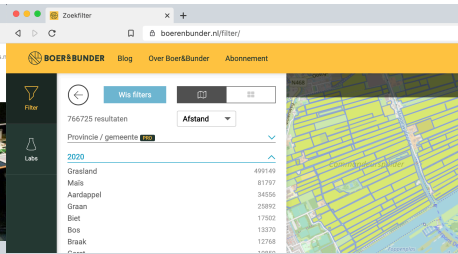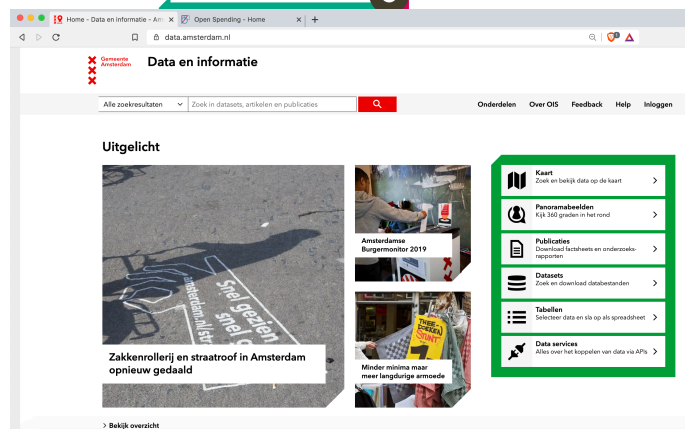
For our research

Statistics Netherlands, open statistical data platform

National open data catalogue data.overheid.nl
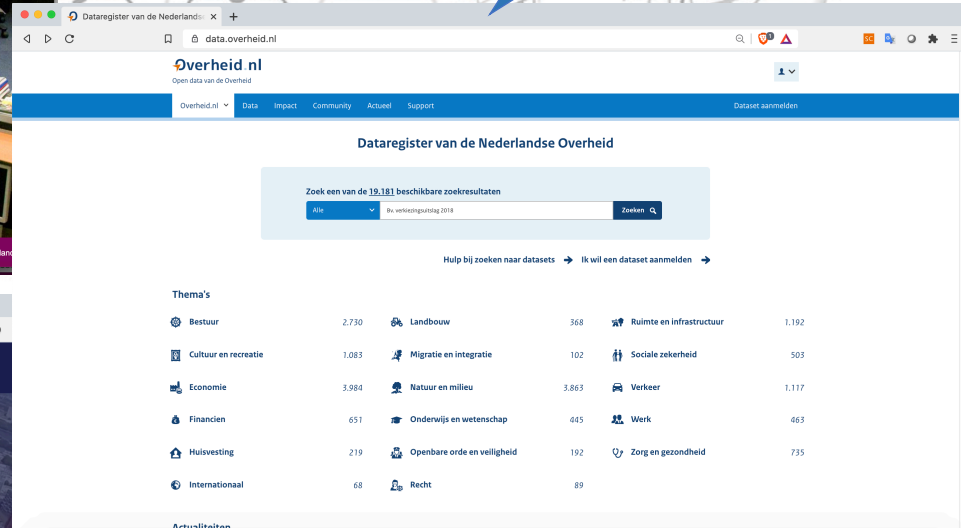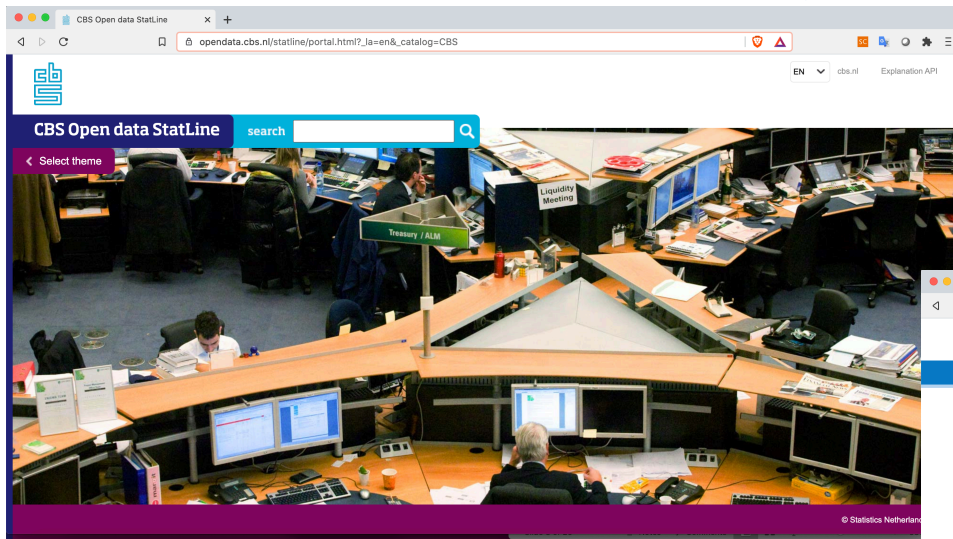
PDOK - Public Service on the Map, national geodata platform

Research questions
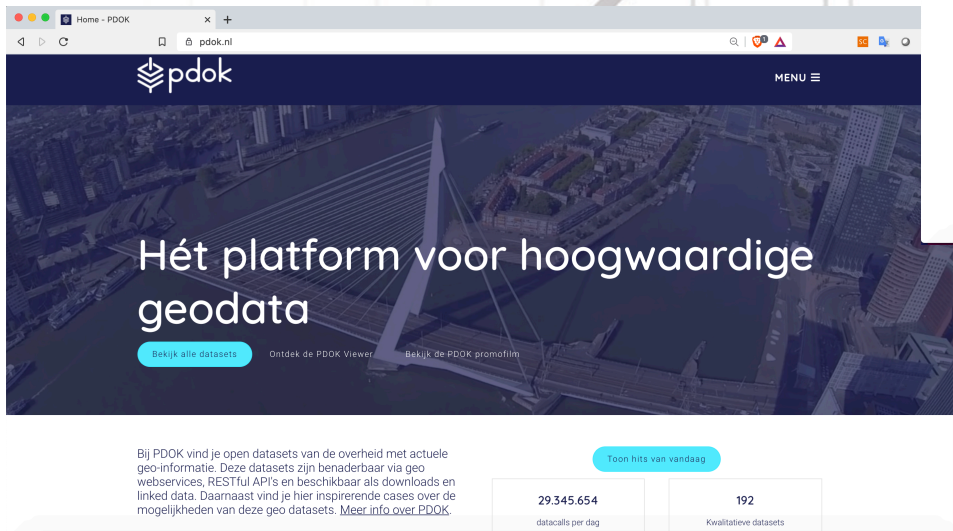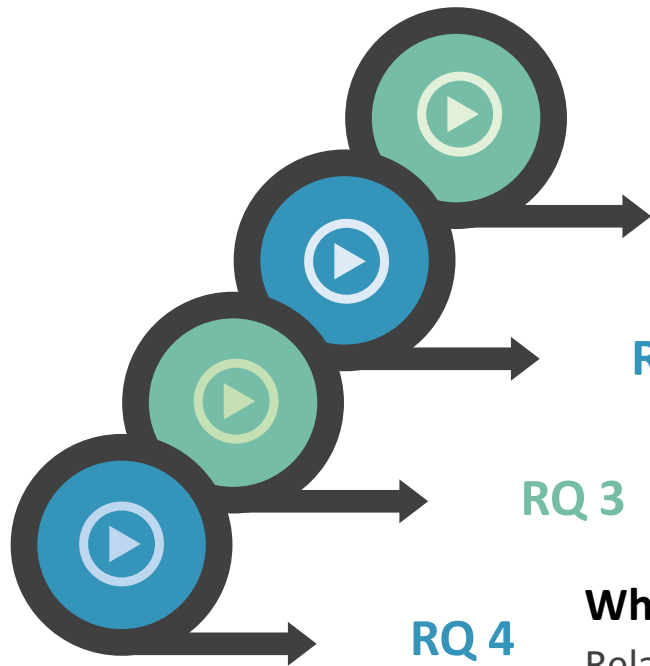
RQ 1

Who are the Open Data users
Types, categories, interrelationships

RQ 2

In which way is open data reused
Purposes, search strategies, access services

RQ 3

Which barriers do open data reusers encounter?
Findability, technical accessibility, data quality

RQ 4

What are the needs of open data reusers?
Related to supply (which data and in which way) and to demand (e.g. user participation)

Research methodologies
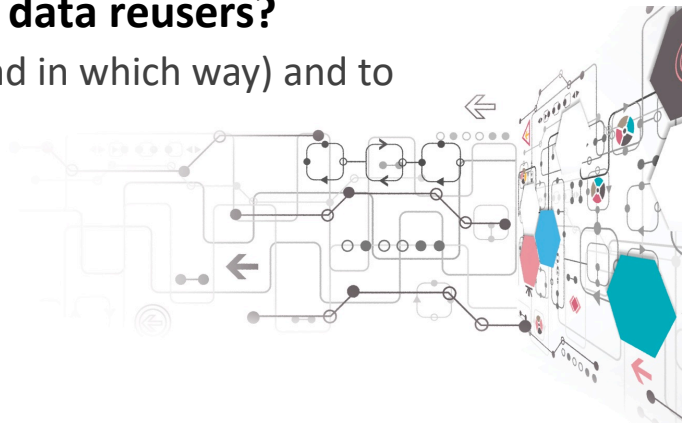
Literature review

- Categories of open data reusers
- Different types of analyses

Quantitative analysis

- IP numbers of log file data Jan-Feb 2019
- Twitter sentiment analysis

Qualitative analysis

- Online survey (Oct '18-Jan '19
- Workshop (March '19)
- Interviews with selected OD users (May '19)
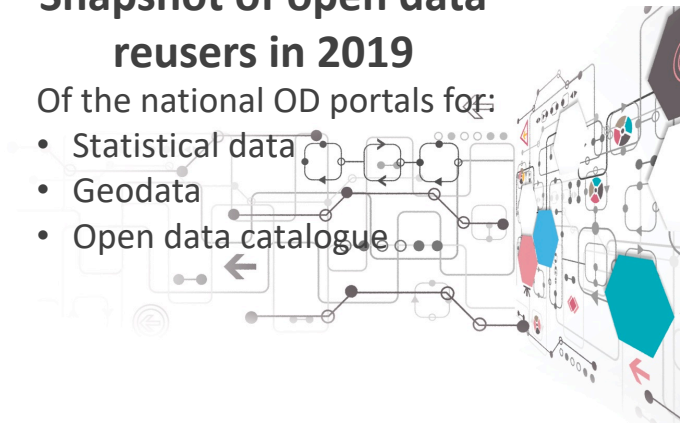
Validation of results

- Quantitative results with suppliers
- Online survey with interviewees (June '19)

Snapshot of open data reusers in 2019

Of the national OD portals for:

- Statistical data
- Geodata
- Open data catalogue

Types of Open Data Reusers

- Commercial
  - Micro enterprises (1-9 staff)
  - Small to medium enterprises (10-249 staff)
  - Large enterprises (≥ 250 staff)

- Not-for-Profit / non-commercial
  - Government organisations
  - NGOs / societal organizations
  - Research / education institutes
  - Concerned citizens

IP Log File Analysis per platform

Twinning Open Data Operational

01 **Top 10 datasets**
Which 10 datasets are requested most often / no. of hits

02 **Which countries**
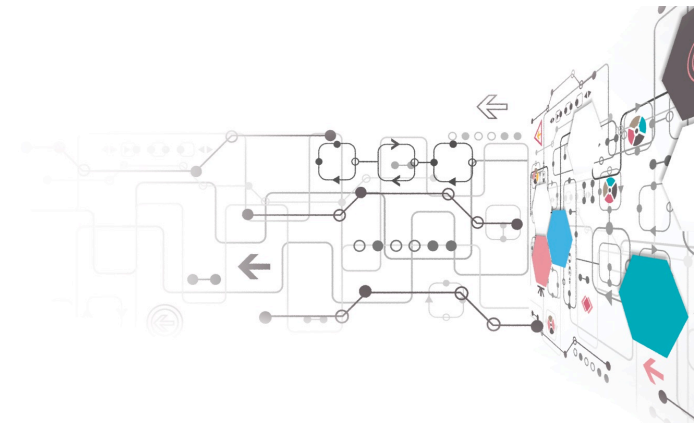Which country do most requests originate from

03 **Top 10 user groups**
Which sectors / types of users can we identify

04 **Top 10 users**
Which large-scale users can we identify

Results Open Statistical Data Portal

- 'Key Figures Neighbourhoods & Districts', and 'Population & Households' most popular datasets

- Relatively many requests from abroad, e.g. USA, Germany and Hong Kong

- Most users in ISP category

- Especially Consumer Prices & Key Figures popular with private sector

- Top 10 large-scale users mainly public sector, higher education and banking/insurance companies
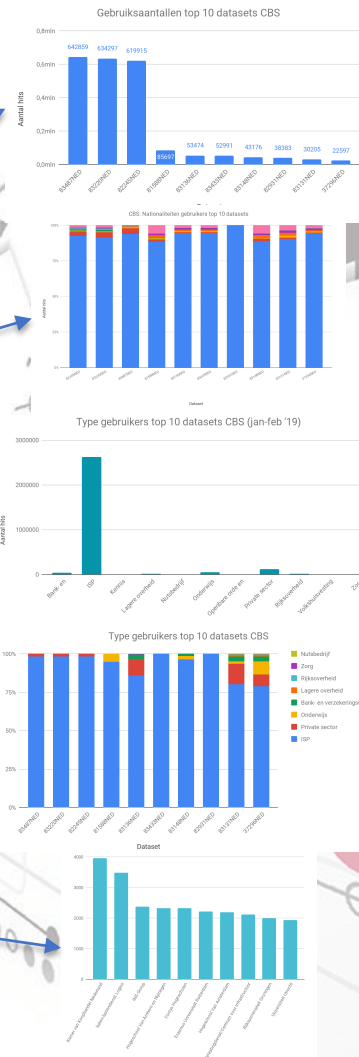
Results PDOK Open Geodata Portal

- Key Registers/Base Data, aerial photography and WOZloket most popular

- Most requests from Netherlands

- Most users ISP category

- Top 10 datasets used by all categories of reusers

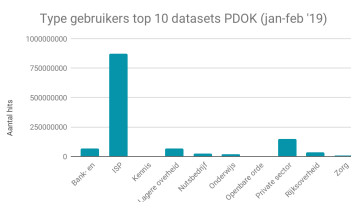- Top 10 large-scale users mainly public sector, higher education and DNB, the Dutch Central Bank
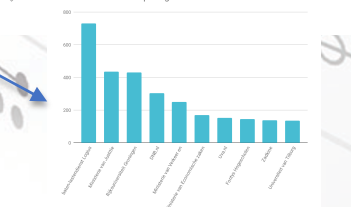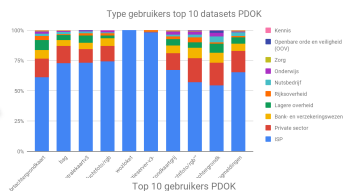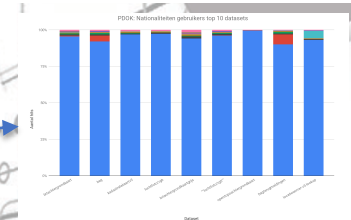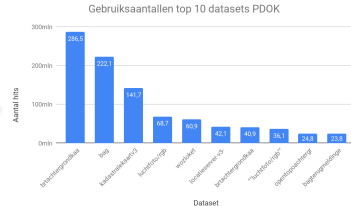
Results data.overheid.nl service
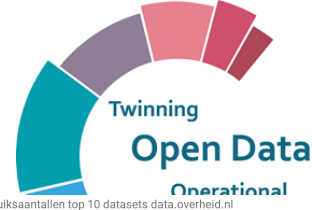
- 'Information Childcare Locations' and 'Accountability Information Housing Associations' most popular

- Most requests from Netherlands, then UK and USA

- Most users in ISP category

- Reasonable spread datasets per category

- Top 10 large-scale users mainly from Ministries, higher education and private sector
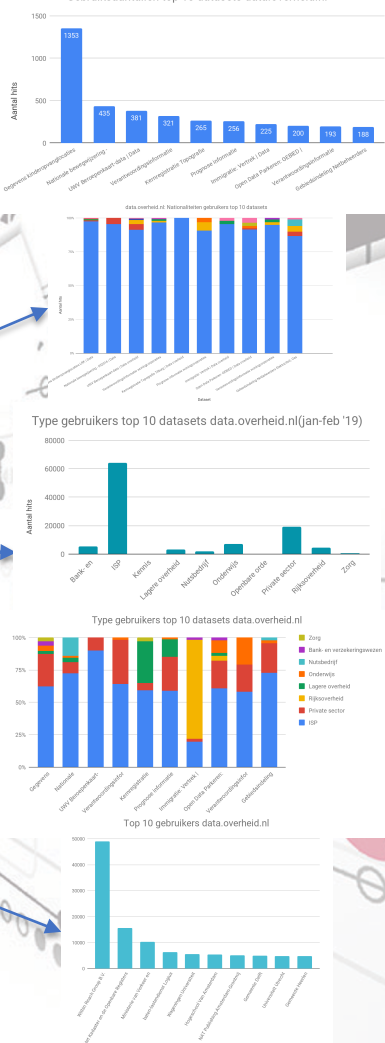
Platform usage large-scale open data users

Platformgebruik grootgebruikers open data
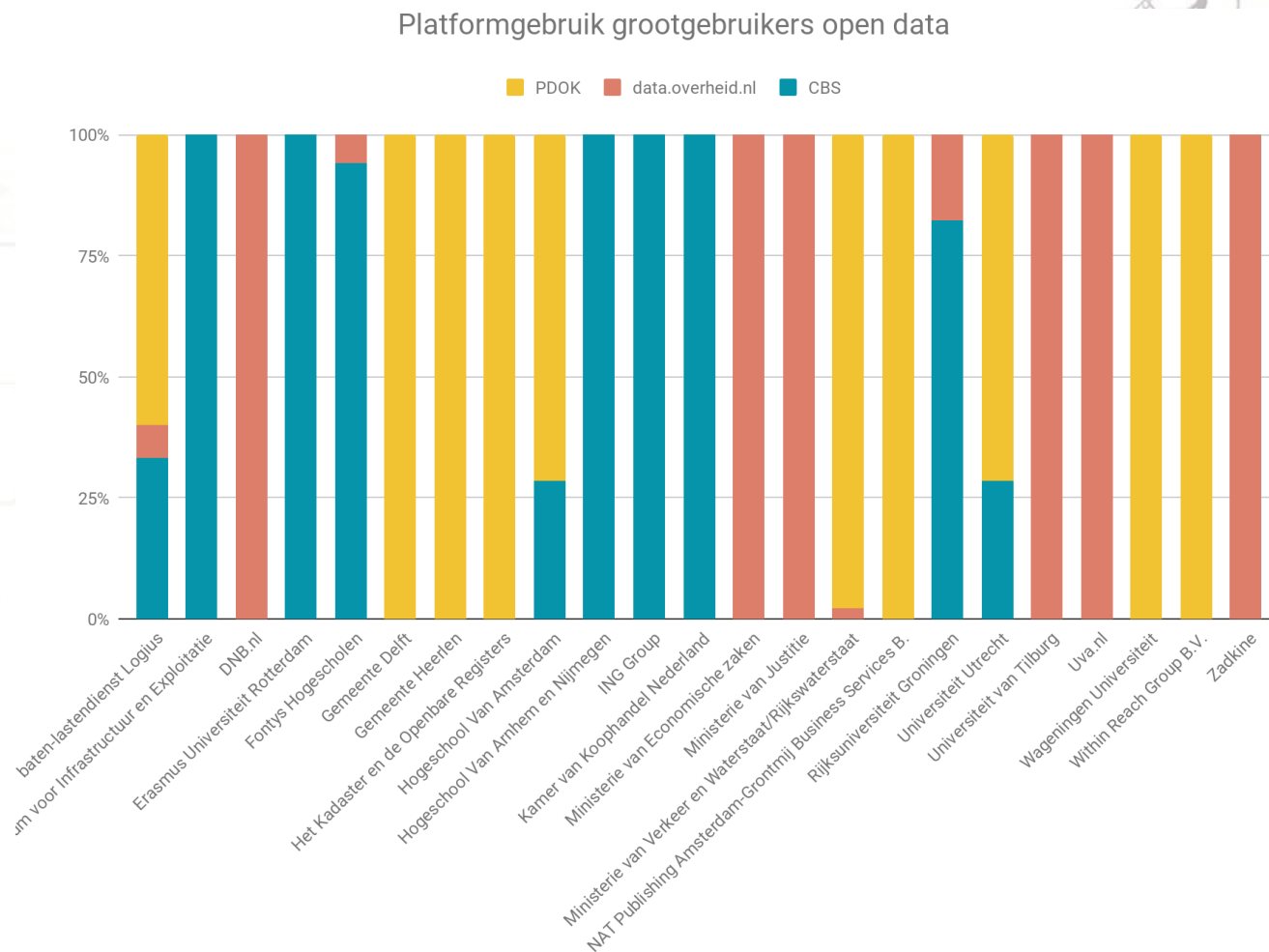
Interviews with OD reusers

- Shortlist of 20 OD reusers, selected from
    - Online survey respondents
    - Internet analysis news items related to open data

- 12 reusers interviewed, with even spread of:
    - Micro enterprises
    - SME / large enterprises
    - NGOs
    - Tinkering citizens
    - Knowledge institute / researchers
    - Public servants

Results interviews

	Commercial organisations	Non-commercial organisations
Commercial purposes	• Fee-based services / apps • Internal business processes • Custom-made products for 3rd parties	• Internal business processes
Societal purposes		• Free services / apps • Community services • As a spin-off of other activities

Twinning
Open Data
Operational

Results interviews

	Commercial organisations	Non-commercial organisations
Commercial purposes	• Fee-based services / apps • Internal business processes • Custom-made products for 3rd parties	• Fee-based services / apps • Internal business processes • Custom-made products for 3rd parties
Societal purposes	• Free services / apps • Community services • As a spin-off of other activities	• Free services / apps • Community services • As a spin-off of other activities

OD search strategies and sources

- Search via
  - Google
  - Own network
  - Social Media
  - Data catalogues (data.overheid.nl, National GeoRegister)
- Sources
  - Mainly direct from original source
  - Via thematic platforms (CBS, PDOK, NDW)
  - Occasionally via commercial enterprises (Esri)

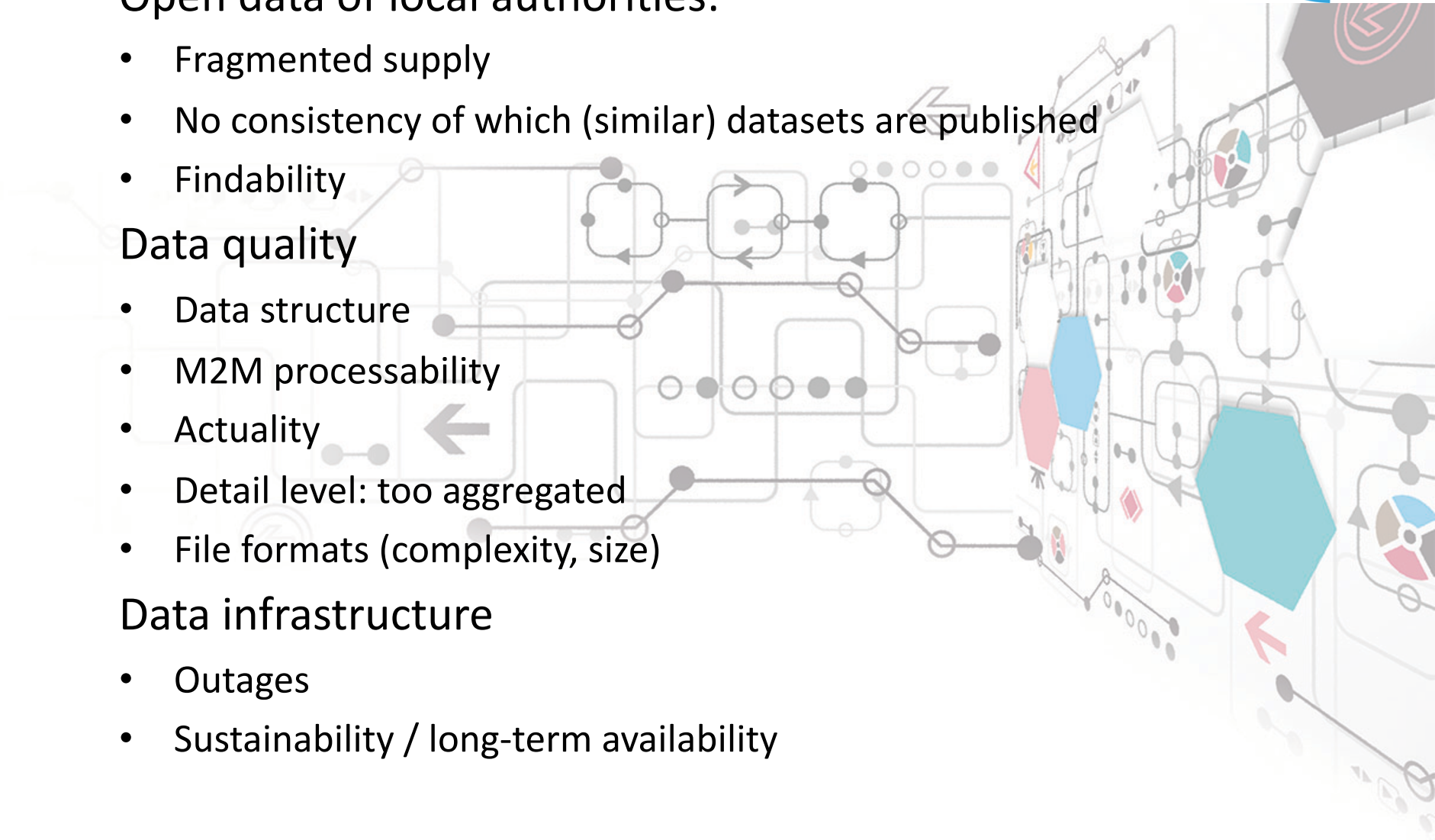# Most mentioned barriers

Open data of local authorities:

- Fragmented supply

- No consistency of which (similar) datasets are published

- Findability

Data quality

- Data structure

- M2M processability

- Actuality

- Detail level: too aggregated

- File formats (complexity, size)

Data infrastructure

- Outages

- Sustainability / long-term availability

Needs of OD reusers

- Improved feedback mechanisms and timely response
- More central direction to coordinate open data of local authorities
- Data of local governments via 1 central platform
- Multiple file formats / multiple services
- More standardisation
  - Data structure
  - Use of international standards for APIs
- Persistent identifiers
- Data governance
  - Governance of open data / open data management
  - CIO/CDO/Data Stewards within the organisation

Suggestions of OD reusers

- Binding agreements with all levels of government and public enterprises which datasets should be available

- Match dates of updates with business deadlines

- Data with higher level of detail made available via APIs

- Provide templates for standardised and consistent coding for similar datasets

- More integration between data platforms

- Government: make better of use expertise of private sector / knowledge institutes to find solutions

- Data Suppliers:
  - Don the cap of reusers more often
  - Visit companies

Main conclusions

- Open data are reused in emerging sectors and for many purposes

- Improve findability of data:
  - intuitive search functionality for non-expert reusers

- Improve open data governance:
  - (Top-down) directing and coordinating open data of local governments and public enterprises
  - Unambiguous semantics / structure / standardisation

- Think in societal challenges and consider which datasets will contribute to solutions
  - involve per challenge stakeholders & scientists
  - Commercial reusers, esp. SMEs have limited time but need to be involved as stakeholders

Future directions for this research

- Need to refine log file analysis

- Need to interview large scale reusers missed in 1st round of interviews

- Need to further analyse cross-links between data platforms, datasets and reusers (which unique IP address used which data, etc.)

- Need to develop tools to systematically analyse reusers

- Need to assess 2nd order effects of open government data (for business cases)

Thank you, and please ask any questions you may have

f.m.welledonker@tudelft.nl