

---

**Big data for policy**  
*from data analytics to evidence-informed policy*

*Kincsö Izsak, Technopolis Group Belgium*

<https://www.data4policy.eu/>

project coordinator: [martijn.poel@technopolis-group.com](mailto:martijn.poel@technopolis-group.com)

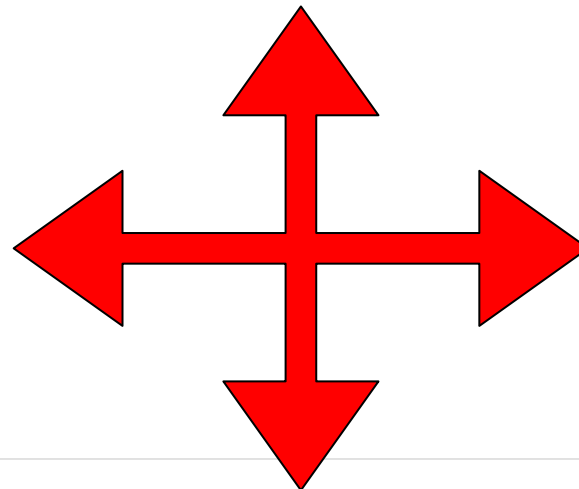
---

## Big data for evidence-informed policy making

*explored the opportunities that innovative data-driven approaches offer for evidence-informed policy making*

Data analytics (theory-based, indicator-based, using statistics, econometrics, modelling, etc.)

Data collection



Visualisation

Policymaking

- Agenda and priority setting
- Policy options
- Policy design
- Ex ante evaluation and impact assessment
- Implementation
- Monitoring
- Ex post evaluation and impact assessment

## Inventory of relevant operational initiatives (58)

---

- Status of the initiative: pilot, demonstrator, implementation
- Type of responsible authority: department, agency, PPP, etc.
- Policy area(s) addressed by the data for policy initiative
- (Expected) use of the data in the policy cycle
- Data sources, variety, volume, velocity
- Data format, data interoperability and data linking
- Data veracity: un/verified, in/complete, in/consistent
- Level of openness of data
- Data analysis methodologies
- Data analytics tools and platforms

**Explorative study,  
non-representative sample**

## Examples related to environmental policy

---



- **Digital Delta in the Netherlands: Public Private Partnership (PPP) on water policy, using sensors, geodata and administrative data**
- **Canadian Geospatial Data Infrastructure (CGDI). Satellite, administrative, sensor and large-scale survey data**

## Examples related to research and innovation



Big data for the evaluation of  
R&D grants in the ICT sector  
(*Spain*)

- Data about R&I policies and impact: online, interactive country reports and cross-country thematic reports
- STAR METRICS repository of datasets and tools to assess the impact of US federal R&D investments
- Data on R&I projects, public procurement, patents, trademarks, publications, websites of ICT companies (e.g. employment), blogs/news sites, job portals

## Social policy, education, transport and other policy areas

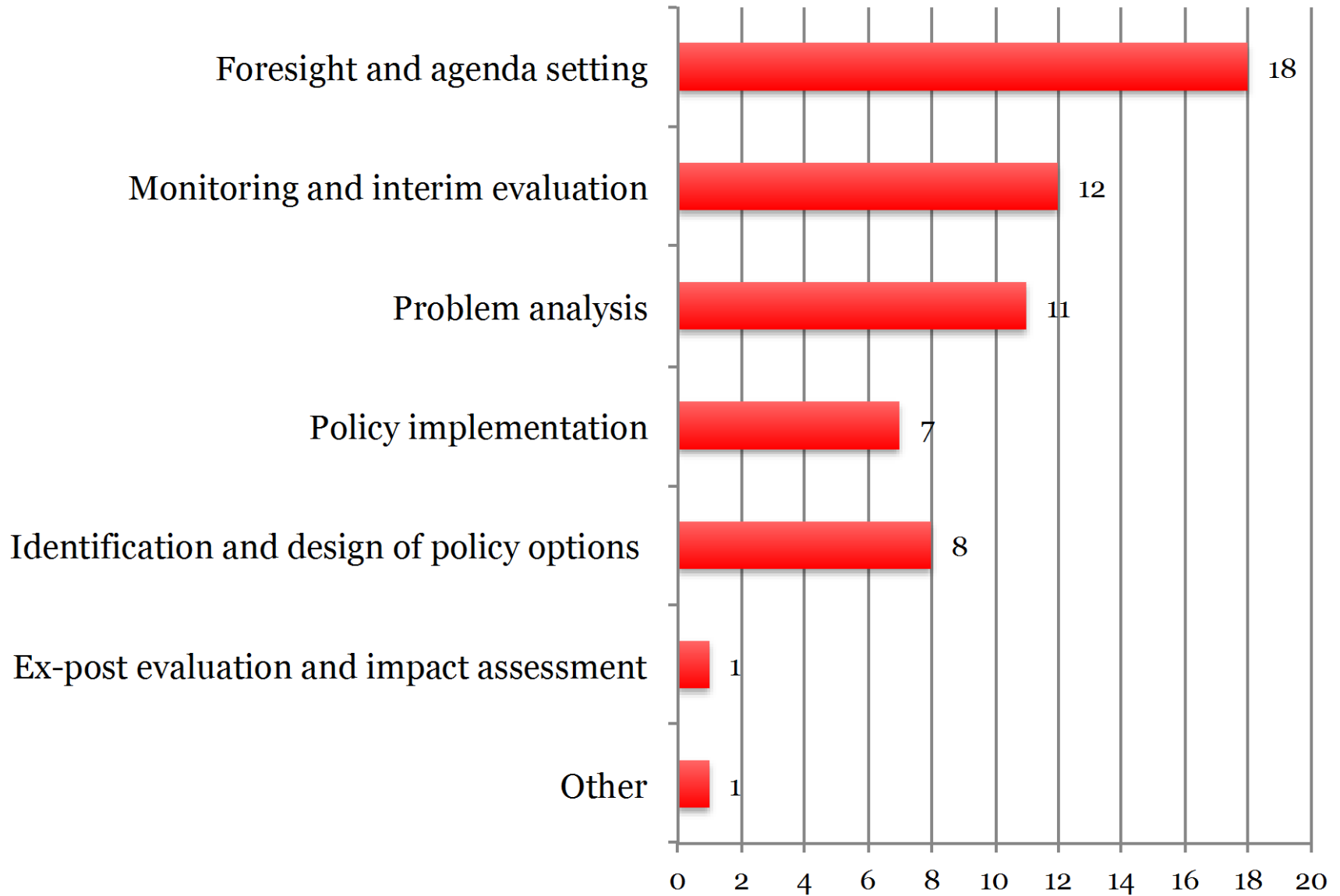


  
GRONINGEN  
DECLARATION

  
Statistics  
Netherlands

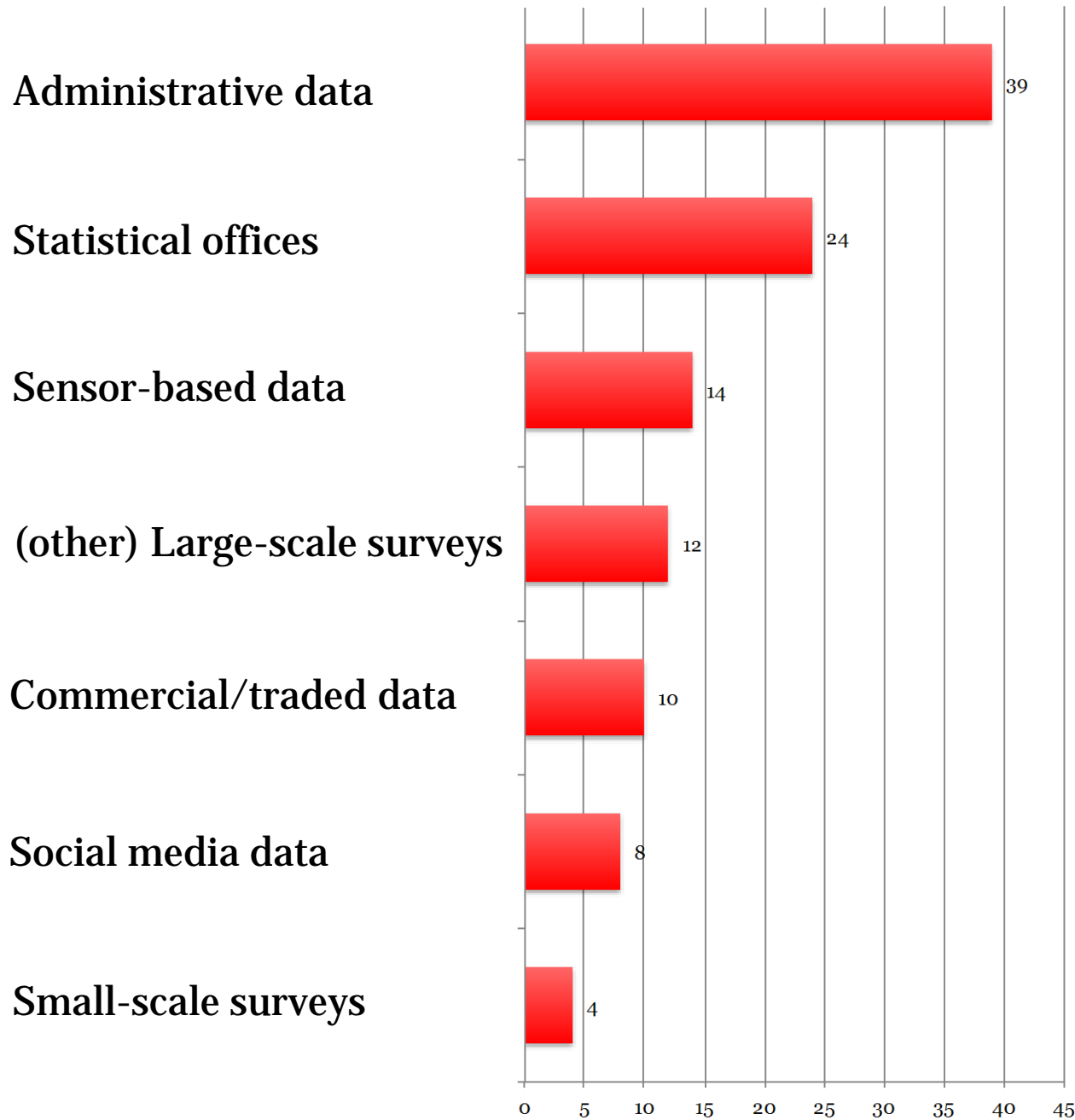
- Free software for open-source micro-simulation of the tax-benefit system in France. Users can calculate tax implications for specific social groups
- Learners control their own education data (from various educational and administrative institutions) and the full dataset allows for analysis of education trends, skills gaps, etc.
- Vehicle detection loops and mobile phone data to monitor and forecast traffic in the Netherlands

# Expected use of data in policy cycle



**n = 58**

## Data sources

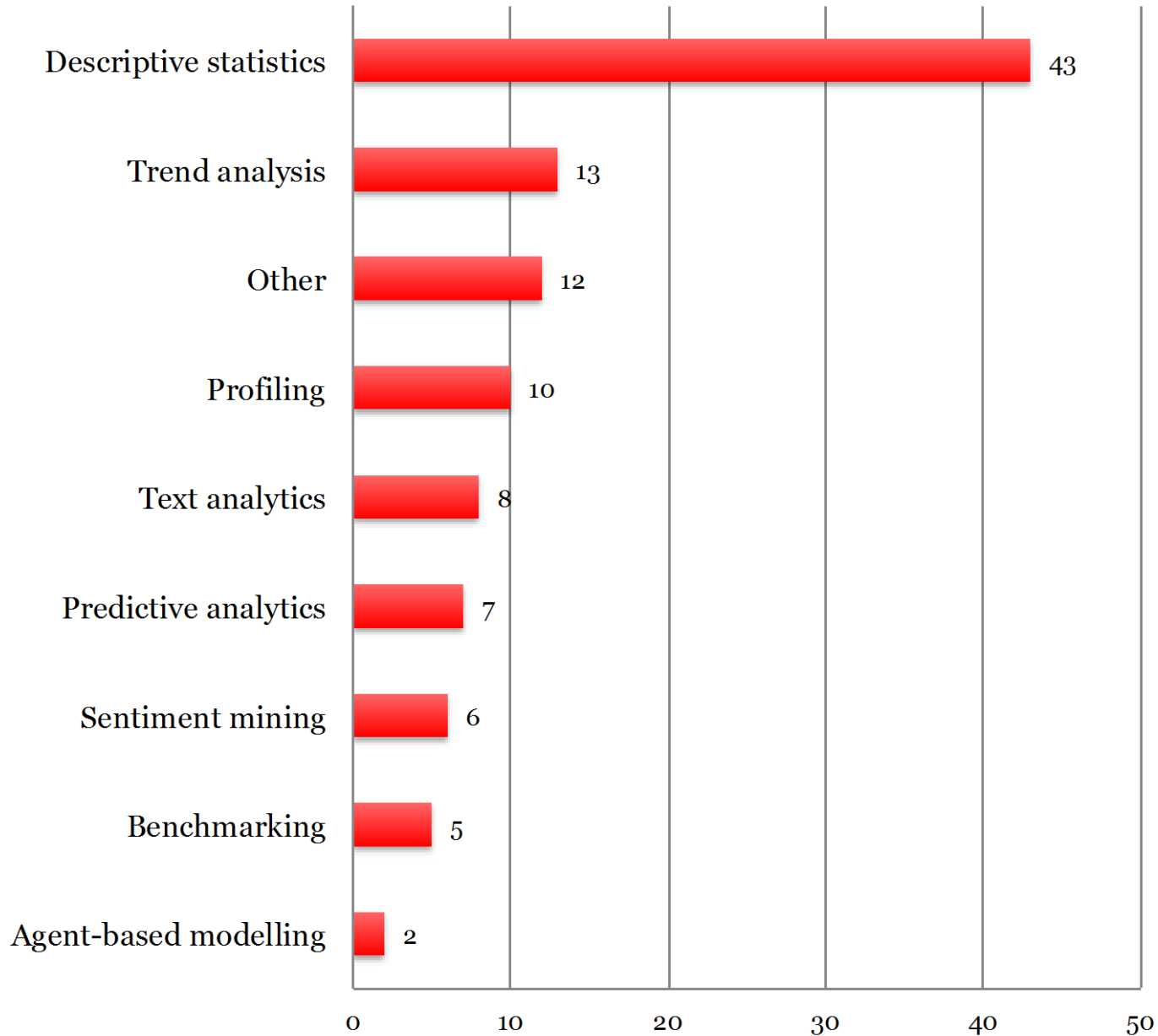


Type of data source closely linked to openness of data: 'half the data open, half the data closed'

58 initiatives  
111 data sources  
Data linking



# Type of data analysis



**n = 58**

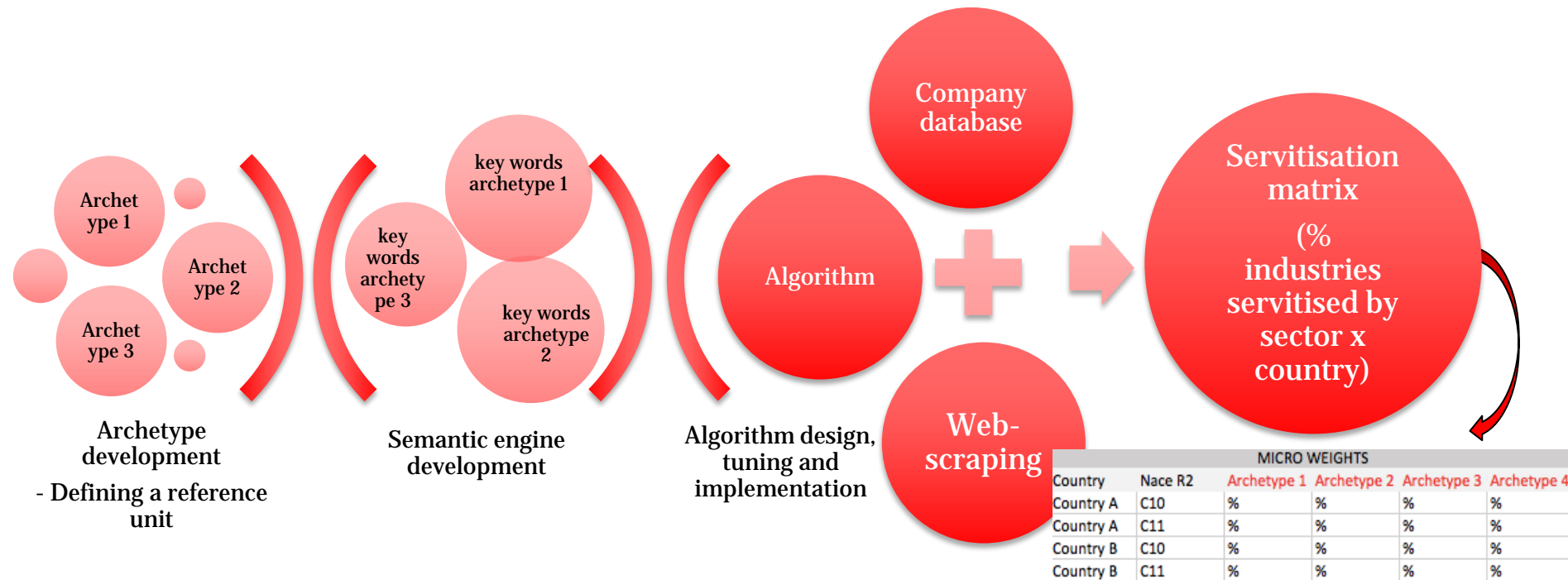
## Main conclusions of the study

---

- Policy for data, e.g. privacy, ownership, standardisation
  - Passing the top of the hype cycle: risk of underestimation of impact
  - Transparency, accuracy, inclusion, accountability (no black boxing)
  - Skills for developing and interpreting data-driven approaches
  - Relevant data or readily available data? Intervention logic!
  - Policy-based evidence: lies, statistics and big data
  - Dual learning: policy experiment + big data experiment
-

## Example of one of our most recent projects: Estimating the level of servitisation through web scraping

- What percentage of firms web-scraped offer product-service combinations as expressed on their websites?
- Using the software Dandelion API (*semantic text analysis as a service*)



## Thank you

---

- [Kincso.izsak@technopolis-group.com](mailto:Kincso.izsak@technopolis-group.com)
- [\*\*Martijn.poel@technopolis-group.com\*\*](mailto:Martijn.poel@technopolis-group.com)