

Gebruik van CBS microdata voor wetenschappelijk onderzoek

Bedrijven, personeel en innovatie

Ruurd Schoonhoven en Marcel van den Berg
Centrum voor Beleidsstatistiek, CBS



**Centraal Bureau
voor de Statistiek**

CBS Microdata Services

- Wat is het?
- Basisregels rond toegang, beveiliging etc
- Wat kun je ermee?
- Bestandencatalogus en koppelen van data
- Organisatie van projecten, kosten
- Een praktijkvoorbeeld
 - Vraagstelling
 - Aanpak
 - Resultaten

Wat is microdata services

- Onderzoek door externe instellingen op CBS microdata
- Gedocumenteerde data uit catalogus evt aangevuld met maatwerk en/of te koppelen (versleutelde) eigen data
- Microdata blijven binnen beveiligde netwerkomgeving
- Benaderen via On Site (werkplek bij CBS) of Remote Access (verbinding vanuit eigen instelling)
- Alleen statistische data als output
- Resultaten moeten gepubliceerd worden
- Kosten

Omvang microdata services

- 120 nieuwe projecten per jaar
- 300 projecten per een jaar actief
- 60 instellingen
- 500 onderzoekers
- > 2000 microdata bestanden in catalogus
- 100 remote access werkplekken
- 600 outputs per jaar

Wie kan toegang krijgen? (CBS-wet)

- Universiteiten, planbureaus, TNO, etc
- Overige instellingen, zoals
 - Onderzoeksafdelingen van overheidsinstellingen
 - Private onderzoeksinstellingen
- Criteria:
 - Alleen voor statistisch of wetenschappelijk onderzoek
 - Instellingen met primaire onderzoeksdoelstelling
 - Publiceren voor de openbaarheid

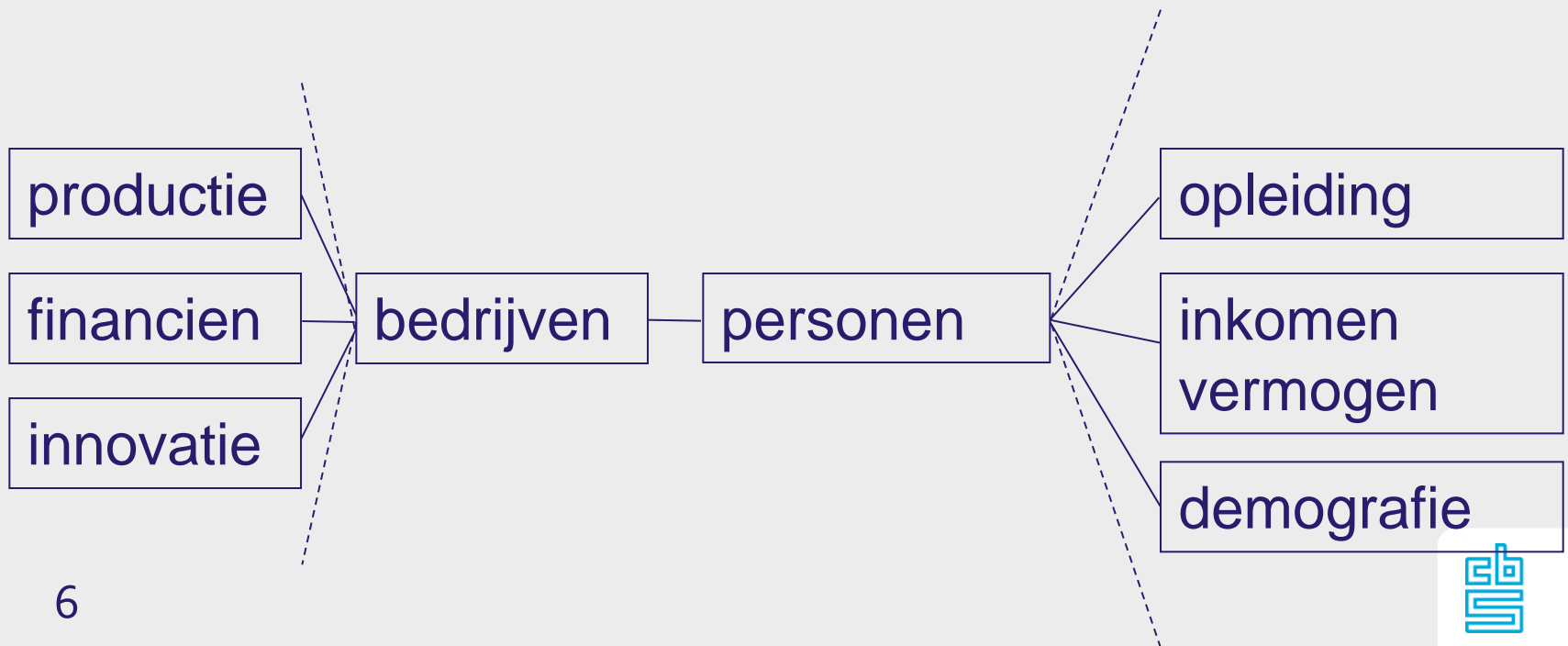
Vertrouwelijkheid van data

....vormt de bestaansgrond van het CBS

- CCS machtiging, contract, geheimhoudingsverklaring
- Aparte RA-PC in afsluitbare ruimte, inloggen met vingerafdruk
- Outputcontrole (geen gegevens van cq herleidbaar tot individuele personen of instellingen)
- Versleuteling van alle identificerende variabelen
- Doelbinding: project krijgt alleen toegang tot die bestanden die nodig zijn om onderzoeksvraag te beantwoorden

Thema overstijgende analyse

- Onderling koppelen van bestanden op
 - Bedrijfs ID
 - Persoons ID
 - ...



Catalogus van microdatabestanden

- Catalogus naar 19 thema's
- Bestanden over plm 200 onderwerpen en 5-20 jaargangen
- Met documentatierapporten op internet
- Bestanden nevenproduct van het statistische proces

Bronnen van de data (1)

- Handelsregister (KvK)
- Registers van de belastingdienst
 - BTW omzet, Vennootschapsbelasting
 - Loonaangifte/polisadministratie (met UWV)
 - Inkomsten- en vermogensbelasting
 - ...
- Sociale zekerheid (WW, AOW, ...)
- Gemeentelijke Basis Administratie
- Basisadministratie Gebouwen (BAG)
- Onderwijsregistraties DUO
- ...

Bronnen van de data (2)

- Eigen CBS enquêtes
 - Financiën van ondernemingen
 - Innovatie en R&D bedrijven
 - Productiestatistiek, Prodcom
 - Internationale handel in goederen
 - ...
 - Enquête beroepsbevolking
 - Gezondheidsenquête
 - Woningbehoefteonderzoek
 - Veiligheidsmonitor
 - ICT gebruik huishoudens
 - 9 - Kiezersonderzoek
 - ...

Bronnen van de data (3)

- Gegevens van andere partijen (soms toestemming nodig)
 - Landelijke Medische Registratie (LMR)
 - Kadaster (woningtransacties)
 - Politie (HALT, HKS)
 - Enquêtes arbeidsomstandigheden (TNO)
 - Kentekenbestanden (RDW)
 - Ziektekosten (Vektis)
 - Medicijngebruik (ZINL, vh CvZ)
 - ...

Statistische eenheden en koppelsleutels

- Personen
 - RIN persoon
- Adressen
 - RIN adres / Object ID (BAG)
- Bedrijven
 - Juridische eenheid (KvK/fiscaal nummer)
 - Bedrijfseenheid (economisch relevant cluster van KvK eenheden) BE_ID
 - Ondernemingsgroep (holdings etc) OG_ID

Fasen van project

- Aanmelding (onderzoeksvoorstel, bestanden)
- Intakegesprek met CBS deskundigen
- Contract, geheimhoudingsverklaring, account
- Evt externe toestemmingen
- Start analyses (met SPSS, Stata, R, ...)
- Outputs
- Publicaties
- Factuur
- Afsluiting met standaard 5 jaar archivering

Kosten (vooraf te begroten)

- Dienstverlening valt buiten centraal CBS budget
- Te dekken: personele lasten plm 15 vte

- Dienstencatalogus met standaardtarieven
- Kosten per project, bepaald door aantal onderzoekers, aantal bestanden, looptijd, extra diensten

- Voorbeelden:
 - Klein project (max 10 bestanden, max 2 onderzoekers, enkele maanden): ~3000 euro
 - Midden project (max 60 bestanden, max 4 onderzoekers, half jaar): ~6500 euro

Contact

- Centrum voor Beleidsstatistiek www.cbs.nl/cvb
 - Microdata services
 - Maatwerkonderzoek
- E-mail cvb@cbs.nl

Een praktijkvoorbeeld

- Promotieonderzoek
 - Promovendus aan de Universiteit Utrecht
 - Leerstoel internationale macro-economie
- Vier verschillende empirische onderzoeken (en één theoretisch onderzoek)
- Tijdsbestek van 3,5 jaar
- Op basis van microdata van (o.a.) het CBS
- Via remote access uitgevoerd

Heel kort: het onderzoek

- “Bevordert internationalisering bedrijfsprestaties?”
- Empirisch onderzoek op CBS-microdatabestanden
- Verschillende dimensies van bedrijfsprestaties:
 - Productiviteit
 - Winstgevendheid
 - Exportsucces
- Verschillende dimensies van internationalisering
 - Importeren (wat, waar vandaan?)
 - Exporteren
 - Exportpromotie

Thema overstijgende analyses

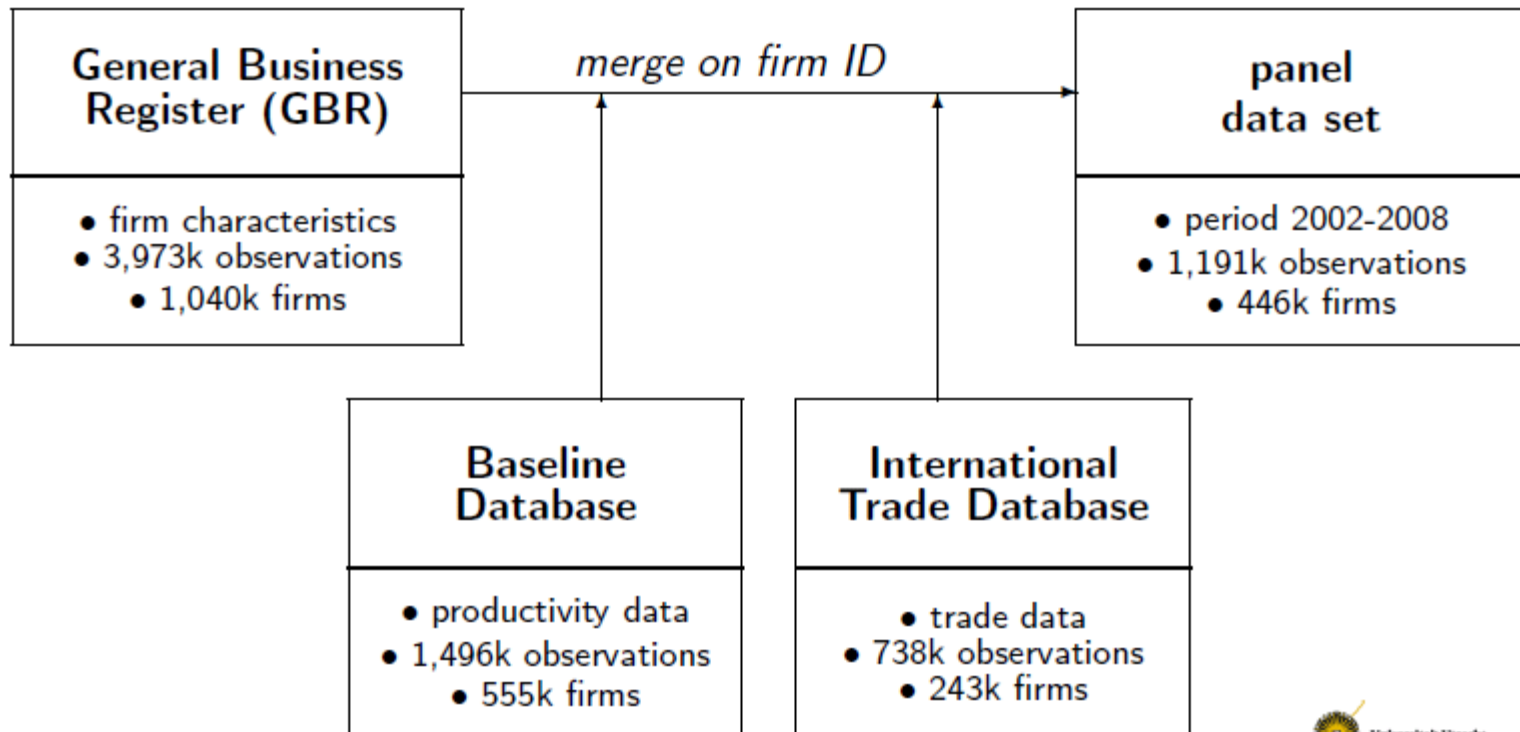
- Dus: koppelen van verschillende microdatabestanden
→ koppelsleutel: bedrijfs-ID
 - Algemeen Bedrijvenregister
 - 'Kapstok' van de bedrijfsregistraties
 - Incl. enkele hoofdkenmerken van bedrijven
 - Sector, omvang, rechtsvorm, locatie,...
 - Economisch statistisch kader
 - Daarmee anders dan bijv. BTW-registratie of handelsregister

Aangekoppelde microdatabestanden

- Internationale Handel in Goederen
 - Invoer, uitvoer (en wederuitvoer)
 - Inclusief verbijzondering naar product en herkomst/bestemming, mits:
 - Buiten EU, of
 - Boven rapportagedrempel
- Financiën van Ondernemingen, BTW-omzet data, etc.
 - O.b.v. bijv. vennootschapsbelasting, BTW-aangifte, inkomstenbelasting
 - Bedrijfseconomische gegevens
 - omzet, (loon)kosten, winst, etc.

Grafische weergave koppeltraject

Figuur: Schematische weergave van koppeling micro-data bestanden



Aankoppelen externe data

- Praktijkvoorbeeld
 - Onderzoeksvraag: “Wat is het effect van exportpromotie op exportprestaties van het MKB?”
 - Evaluatie van exportpromotie-instrument van RVO.nl
 - Deelnamebestanden op KvK-nummer
 - Importeren in RA-omgeving CBS
 - Koppeltabellen KvK-nummer - fiscaal nummer - bedrijfs-ID
 - Resultaat: microdatabestand waarin exportprestaties van deelnemende bedrijven kunnen worden gevolgd in de tijd.

Heel kort: onderzoeksresultaten

- Bedrijven die importeren zijn productiever dan bedrijven die dat niet doen
- De causaliteit loopt van productiviteit naar importeren, niet vice versa
- De hoogte van de productiviteitspremie wordt beïnvloed door de herkomst van en het type product dat wordt geïmporteerd
- Bedrijven die exporteren realiseren geen hogere winstmarges dan bedrijven die dat niet doen
- Bedrijven die een begeleide exportstart maken realiseren geen hogere exportgroei dan bedrijven die zonder hulp gaan exporteren, maar hebben wel een grotere kans om op de langere termijn actief te blijven als exporteur

Scala aan mogelijkheden

- Dus, bedrijven vormen de kapstok (ABR)
 - wat hang je er verder aan?
 - Bedrijfseconomische gegevens (PS, BTW, SFGO, SFKO, etc.)
 - Internationale handel (IHG)

Maar ook mogelijk:

- Innovatie (CIS/R&D-enquêtes)
 - Actueel voorbeeld: kruisen handel en innovatie
- Werknemers en kenmerken van werknemers (SSB)
 - 'employer-employee linked data'
- Zelfstandigen/ ZZP'ers met kenmerken
-?