



Digitalisering: Over de waarde van data voor de samenleving

Het Centraal Bureau voor de Statistiek heeft een unieke infrastructuur en positie om de overheid en de samenleving te voorzien van hoogwaardige en betrouwbare statistiek voor beleid, uitvoering en wetenschap. Het heeft de afgelopen jaren veel kennis en ervaring opgedaan op het gebied van het verzamelen van data, datakoppeling, data-analyse en datagedreven werken. Daarbij worden nieuwe technologieën, waaronder AI-toepassingen en datavirtualisatie ontwikkeld met innovatieve (inter)nationale partners. Door de overheid wordt in het kader van het datagedreven werken bij de uitvoering van overheidstaken een steeds groter beroep gedaan op het CBS.

Visie op datagebruik

Het gebruik van het aantal databronnen en de hoeveelheid data nemen in hoog tempo toe, evenals de mogelijkheden om deze data te gebruiken. Een heldere visie op de kansen en risico's van data is onmisbaar. De impact op de maatschappij is immers groot.

De overheid moet een stevige eigen positie blijven behouden in de datarevolutie die door bedrijven zoals Google, Facebook en Amazon wordt ontketend. Het CBS is van oudsher het (big) data bedrijf van de overheid en is gepositioneerd voor deze taak, mede om ervoor te zorgen dat de overheid voor de uitvoering van publieke taken niet geheel afhankelijk wordt van grote commerciële bedrijven, maar ook zelf over de kennis, de data en de infra beschikt.

Bij datagebruik voor en door de overheid staan belangrijke waarden centraal. Informatie moet betrouwbaar, kwalitatief hoogwaardig en onafhankelijk zijn. Deze wordt immers gebruikt voor het maatschappelijk debat, onderzoek, beleidsontwikkeling en besluitvorming. Het vinden van de juiste data en betrouwbare databronnen en het kunnen interpreteren van deze data naar beleid is daarbij nog een grote uitdaging. Maar het maatschappelijk effect kan groot zijn.

Een voorbeeld uit de praktijk: de gemeente Den Haag heeft gericht beleid ontwikkeld voor de verlichting van armoede onder kinderen. Op basis van diverse databestanden wordt meer duidelijk over de profielen van gezinnen en de wijken waarin die kinderen leven. Zo blijkt dat ook in gezinnen waar ouders beiden werkzaam zijn, toch ook kinderen met risico op armoede zitten. Deze gezinnen wonen bovendien in andere wijken dan die in eerste instantie in het

vizier waren voor armoedeproblematiek. Hierdoor kan het gemeentelijke armoedebeleid effectiever worden uitgevoerd.

Issues rondom het gebruik van data

Het invoeren van datagedreven werken vereist zorgvuldigheid. Issues rondom waardevol gebruik van data die spelen zijn:

- het omgaan met de versnippering van het datalandschap
- het borgen van privacy en informatieveiligheid
- het zorgdragen voor navolgbare en eerlijke algoritmes
- het waarborgen van regie bij de burger
- de gevolgen van steeds meer data bij private partijen.

1. Versnippering van het datalandschap: Afspraken en standaardisering

Datagebruik door de overheid vraagt om afspraken over het standaardiseren van data zodat data ook te koppelen en uitwisselbaar zijn. Op dit moment is het datalandschap sterk versnipperd. Als iedereen data op zijn eigen manier metadataert, opslaat en bewerkt wordt koppeling en correct gebruik ernstig bemoeilijkt en het risico op foutieve koppeling en interpretatie vergroot. Een gemeenschappelijke data-architectuur en afspraken over metadatering en standaardisering zijn noodzakelijk. In een aantal landen, zoals Estland en Nieuw-Zeeland is hier veel sterkere nationale regie op, onder andere door het aanstellen van een nationale chief data officer of chief data steward.

2. Nieuwe technologie voor datakoppeling met behoud van privacy

Met een toename van data en datagebruik wordt het van nog groter belang om privacy en informatieveiligheid te borgen. Nieuwe technologieën zoals datavirtualisatie en secure multiparting computing maken het mogelijk om

- databestanden te behouden op de plaats waar ze zijn opgeslagen
- op afstand te koppelen
- databestanden te koppelen van meerdere partijen zonder dat de ene partij in de bestanden kan of de bestanden ziet van de ander
- en op geaggregeerd niveau (dus nooit op het niveau van de individu) wel te beschikken over de eindresultaten

Op dit moment zijn er al experimenten met onder meer CBS, RUG, Maastricht University en andere organisaties die laten zien dat deze technologieën in de praktijk toepasbaar zijn. Het is nodig om te investeren in dit soort nieuwe technologieën die goed gebruik van data mogelijk maken waarbij privacy en informatieveiligheid geborgd zijn.

3. Eerlijke en transparante algoritmes

Het maken van overheidsbeleid met data waarvan de afkomst niet helder is en met methodieken cq algoritmes die een black box zijn, brengt risico's met zich mee. Hoe komen databestanden tot stand en hoe worden deze verwerkt? Het is

van cruciaal belang om afspraken te maken over openheid rondom algoritmes en methodieken om de betrouwbaarheid en de navolgbaarheid daarvan te kunnen borgen. Daarbij gaat het met name om de data waarop de algoritmen gebaseerd zijn.

4. De burger heeft recht op inzage

Het uitgangspunt voor een fatsoenlijk databeheer is waar mogelijk de regie op de eigen data bij burgers te leggen. De burger moet zoveel mogelijk zelf kunnen bepalen wie zijn persoonsgegevens mag gebruiken. In Estland is dit het uitgangspunt. Daar is goed geregeld wie toegang heeft tot welke persoonsdata. Technologie kan ervoor zorgdragen dat een burger te allen tijde kan zien welke instantie en welke persoon in die instantie iets doet met jouw data en met welk doel. Een transparante overheid kan aangeven welke data waarvoor worden gebruikt.

5. Nieuwe databronnen

Steeds meer data zijn en komen in handen van private aanbieders zoals technologiebedrijven. Zij worden hiermee ook een bron van data voor derde partijen, waaronder overheden. Om betrouwbare gegevens te kunnen borgen zijn afspraken nodig over het verstrekken van private databronnen aan de overheid. Daarbij zijn doelbinding, data minimalisatie en proportionaliteit relevant.

Voorbeeld: datasysteem Estland

De manier van werken in Estland kan een voorbeeld voor Nederland zijn. In Estland worden administratieve data en data van de private sector (virtueel) bijeen gebracht. Door een gemeenschappelijke data-architectuur, metadatering en standaarden kunnen voor administratieve toepassing gegevens worden gebruikt. De burger heeft inzage in welke organisatie welke data gebruikt en waarvoor. Overheidsdata en data van bedrijfsleven worden daarnaast gebruikt voor statistiek en beleidsanalyse op geaggregeerd niveau. Er worden dus geen individuele gegevens publiek.

Door meer regie op en investeringen in data kunnen kansen worden benut en risico's gemitigeerd.

Voor meer informatie: Arjan de Jong,
Beleidsadviseur | Staf DG | Contactpersoon CBS - Tweede Kamer
T 070 337 4431 | M 06 5084 2616 | ag.dejong@cbs.nl