



Kwaliteitscontrole externe bronbestanden en de AVG/CBS-wet

5.1.2.e

FG van het CBS

projectnummer Klik en typ projectnummer
 Klik en typ sector
 20 februari 2020

samenvatting Klik en typ de samenvatting
trefwoorden Klik en typ de trefwoorden

1. Inleiding

In een eerder advies¹ van de FG is aangegeven dat onder voorwaarden microdata terug geleverd kan worden aan bronhouders in het kader van aanvullende statistische diensten (ASD).

In het meer reguliere proces worden er bronnen als input voor het reguliere statistiekproces ontvangen van bronhouders (leveranciers)². Deze data wordt bij binnenkomst gecontroleerd op bruikbaarheid. De vraag is of en hoe over de kwaliteit van deze databestanden aan de leverancier gecommuniceerd gaat worden. Een bijzonder geval van een kwaliteitsissue is waar in een later stadium als gevolg van vragen aan het CBS onduidelijkheid ontstaat over de correctheid/compleetheid/interpretatie van de ontvangen data. Men mag hopen dat dergelijke laat opgemerkte kwaliteitsissues een ad hoc karakter hebben, welke ook een ad hoc aanpak zullen vereisen. De hieronder voorgestelde aanbevelingen zouden hierbij wel als leidraad kunnen dienen. Bij het terug leveren van microdata als voorbeeld waar kwaliteitsissues zijn, gelden eigenlijk dezelfde eisen als bij het leveren van microdata bij een ASD.

Er zijn een aantal opties welke in het licht van de kaders, in dit geval de CBS-wet en de AVG, in meer of mindere mate een voorkeur hebben.

¹ Basisprincipes bij verwerkingen van microgegevens bij werk voor derden.

² Deze nota gaat over registerhouders en niet over individuele berichtgevers

2. Uitgangspunten CBS-wet en AVG

Heel in het kort een paar kaders en uitgangspunten vanuit de CBS-wet en de AVG op basis waarvan keuzes gemaakt kunnen worden of en hoe kwaliteitsissues gemeld worden aan de leverancier.

Heel in het kort vraagt de AVG in dit kader:

1. De verwerkingen binnen de betreffende grondslag en de doelbinding uit te voeren;
2. Het aantal verwerkingen van persoonsgegevens te minimaliseren;
3. Het aantal te verwerken persoonsgegevens te minimaliseren (minder records, minder variabelen);
4. De verwerkingen zo veilig mogelijk uit te voeren.

Ook heel in het kort vraagt de CBS-wet in dit kader:

Artikel 37

1. De door de directeur-generaal in het kader van de uitoefening van de taken ter uitvoering van deze wet ontvangen gegevens worden uitsluitend gebruikt voor statistische doeleinden.
2. De in het eerste lid bedoelde gegevens worden niet verstrekt aan anderen dan degenen die belast zijn met de uitvoering van de taak van het CBS.
3. De in het eerste lid bedoelde gegevens worden slechts zodanig openbaar gemaakt dat daaraan geen herkenbare gegevens over een afzonderlijk persoon, huishouden, onderneming of instelling kunnen worden ontleend, tenzij, ingeval het gegevens met betrekking tot een onderneming of instelling betreft, er een gegronde reden is om aan te nemen dat bij de betrokken onderneming of instelling geen bedenkingen bestaan tegen de openbaarmaking.

3. De overwegingen en aanbevelingen

Om de taak van het CBS uit te voeren heeft het CBS data nodig van een bepaalde kwaliteit. Deze data wordt opgevraagd bij bronhouders. Met de bronhouder worden o.a. afspraken gemaakt over inhoud, frequentie, kwaliteit en specificaties inclusief meta-informatie van deze leveringen.

De bronhouder controleert bij levering zelf de kwaliteit op vooraf vastgestelde checks. Dit kunnen bijvoorbeeld waardenchecks zijn maar ook volledigheidchecks (aantal records).

Idealiter geldt:

1. de leverancier weet welke kwaliteit geleverd moet worden.
2. het CBS weet welke controles de leverancier heeft uitgevoerd bij levering.
3. het CBS weet de uitkomst van de controles bij levering.
4. De leverancier levert conform specificaties.

Het CBS kan aanvullende controles doen maar het is aan te bevelen dat de leverancier zelf de controles uitvoert welke deze ook zelf kan uitvoeren. Dit maakt de zaken ook AVG-technisch meer zuiver. Het gevaar bestaat namelijk dat als het CBS werk (=controles) van de leverancier uitvoert de uitkomsten van dit werk ook voor andere doeleinden dan statistiek en wetenschap gebruikt gaan

worden door deze leverancier. Bij de Belastingdienst zou dit bijvoorbeeld kunnen leiden tot naheffingen bij berichtgevers. Daar wil het CBS buiten blijven. Dus ja, de Belastingdienst had dit zelf kunnen doen. En ja, een beter bronbestand is beter voor de statistiek (CBS). Maar liever dus dat de Belastingdienst het zelf doet. Het is tenslotte ook hun taak.

Aanbeveling 1: Het CBS streeft ernaar dat de leveranciers zelf de controles uitvoeren welke zij zelf kunnen uitvoeren.

Indien de leverancier niet levert conform de specificaties dan kan het CBS de leverancier hier op wijzen zodat het leidt tot een nieuwe levering en/of leidt tot goede vervolgleveringen in de toekomst. Probeer daarom de specificaties zo helder mogelijk te formuleren. Dan kan ook in termen van deze specificaties gecommuniceerd worden. Dit voorkomt dat er op niveau van individuele records gecommuniceerd hoeft te worden met de leverancier.

Aanbeveling 2: Beschrijf de specificaties (kwaliteit) van de gewenste dataset zo gedetailleerd mogelijk.

Aanbeveling 3: Streef er naar dat er niet op individueel recordniveau gecommuniceerd hoeft te worden met de leverancier. Communiceer bij voorkeur in termen van de reeds geformuleerde specificaties.

Bovenstaande aanbevelingen hebben het doel óf nu een goede herleverantie te krijgen en/óf in de toekomst wel goede leveranties te krijgen. Dit geheel conform de PDCA. Er is mogelijk ook een andere optie: indien je voldoende informatie hebt kun je zelf het bestand aanpassen zodat de kwaliteit voldoende is voor verdere verwerking. Dit kan een mogelijkheid zijn indien de leverancier (nog) niet bereid is haar eigen proces aan te passen en bekend is welk probleem er in de data zit. Terugkoppeling zoals eerder omschreven heeft natuurlijk alleen zin indien de leverancier 'beterschap beloofd' en/of bereid is om het probleem te onderzoeken. Indien de statistiekproductie echt in gevaar komt dan zal er een ad hoc procedure gestart moeten worden.

Aanbeveling 4: Als de leverancier niets doet met de terugkoppeling dan is het beter helemaal niet terug te koppelen.

Als blijkt dat dataset onvolkomenheden bevatten welke niet eerder voorzien waren en dus nog niet opgenomen zijn in de specificaties dan zal in overleg met de leverancier gekeken moeten worden of deze specificaties (en de daarbij behorende controles) aangepast moeten worden. Dit is de PDCA op de specificaties en controles. Als het CBS achter deze onvolkomenheden komt dan probeert het CBS de leverancier te overtuigen haar controles aan te passen. Het CBS kan deze aanvullende controles formuleren in tekst en/of scripts en/of individuele voorbeelden waar het fout gaat. Waarbij de laatste optie de minst gewenste optie is.

Aanbeveling 5: Communicatie met de leverancier geschiedt bij voorkeur zonder vermelding van individuele records, voorbeelden of scripts.

Het bovenstaande raakt de kern van ons bestaan, namelijk het vertrouwen dat aan het CBS verstrekte gegevens niet verspreidt worden. Met name bij secundaire bronnen is het terugleveren van info, zelfs alleen algoritmes, vaak al onthullend. Immers de bronhouder heeft de mogelijkheid om dezelfde algoritmes uit te voeren en vervolgens de lijst met individuele records te krijgen. Dit kan gevolgen hebben voor betrokkenen en is dus hoogst ongewenst.

Omgevingssensitiviteit is vereist en het is zaak een tunnelvisie uit te sluiten. Een teruglevering is altijd maatwerk en dient zorgvuldig afgewogen te worden. Het volgen van de aanbevelingen is een goede zaak, maar is niet voldoende om fouten te voorkomen. Een peer-review is aanbevelenswaardig.

Regel 1: Voor dat je een eenmalige, of eerste, terugkoppeling naar een leverancier doet:

- Overweeg of je het acceptabel zou vinden als het je eigen gegevens betrof;
- Bespreek de situatie, waarom en wat je terug wilt leveren met een collega die op de hoogte is van de materie;
- Mocht je met je collega van mening verschillen leg het dan neer bij je manager en/of vraag advies van een kwaliteitsmanager cq privacy ambtenaar;
- Leg het incident én de oplossing (zeker ingeval van terugleveren van individuele data) vast in de incidentendatabase (ISO9001).

Even nog voor de volledigheid: fouten in leveranties worden alleen terug gecommuniceerd indien de leverancier hier ook iets aan kan (wil) doen. Natuurlijk kunnen leveranties fouten bevatten indien deze geconfronteerd worden met datasets uit bronnen van andere leveranciers. Over dit soort fouten wordt nooit op individueel niveau gecommuniceerd met de leverancier. Natuurlijk kan wel in de vorm van een statistiek informatie worden over de kwaliteit van bepaalde bronnen teruggekoppeld worden. Maar dan is het statistiek en zal dit ook gepubliceerd moeten worden.

Regel 2: met een leverancier wordt nooit op individueel recordniveau gecommuniceerd over kwaliteitsissues welke de leverancier niet zelf kan oplossen.

Versiehistorie en distributie

Versie	Datum	Beschrijving van de wijziging	Auteur
1.0.p1	Augustus 2019	Initiële versie	MBLN
1.0.p2	Februari 2020	Wijzigingen n.a.v. eerste consultatie	MBLN
	dd-mm-jj	Geplande update.	

Versie	Distributie
1.0.p1	5.1.2.e CBO-J
1.0.p2	DG, Hoofddirecteuren, CBO-J