



PBLQ

**DUURZAAM DIGITAAL DATABEHEER
BIJ DE RIJKSOVERHEID: EEN VERKENNING**

Eindrapport
versie 1.0
26 februari 2021

Inhoudsopgave

Managementsamenvatting	2
1. Inleiding	4
1.1 Aanleiding	4
1.2 Opdrachtformulering en scope	5
1.3 Werkwijze	5
1.4 Leeswijzer	6
2. Definities en afbakening	7
2.1 Inleiding	7
2.2 Definities	7
2.3 Afbakening van deze verkenning	8
3. Juridische en ethische kaders voor duurzame toegankelijkheid van data	10
3.1 Inleiding	10
3.2 Uitdagingen	10
3.3 Normen en kaders	10
3.3.1 Wet openbaarheid van bestuur	11
3.3.2 Wet hergebruik van overheidsinformatie	11
3.3.3 AVG	12
3.3.4 Archiefwet	14
3.3.5 Ethische overwegingen	15
3.4 Samenvatting	16
4. Organisatorische aspecten	17
4.1 Inleiding	17
4.2 Uitdagingen	17
4.3 Normen en kaders	17
4.3.1 Datamanagement	17
4.3.2 Datagovernance	18
4.3.3 Gegevensmanagementrollen	19
4.3.4 De rol van de reguliere medewerker	20
4.4 Samenvatting	20

5.	Technische aspecten	21
5.1	Inleiding	21
5.2	Uitdagingen	21
5.3	Normen en kaders	22
5.3.1	DUTO-eisen en <i>archiving by design</i>	22
5.3.2	FAIR en NORA gegevensmanagement	23
5.3.3	NEN(-ISO) normen	23
5.3.4	Metadatamodellen	23
5.3.5	Instrumenten voor waardering en selectie	23
5.3.6	Algoritmes	24
5.3.7	Overige kaders	24
5.4	Samenvatting	25
6.	Casus: databeheer bij het CBS	26
6.1.1	Inleiding	26
6.1.2	Organisatie van het databeheer	26
6.1.3	<i>Lessons learned</i>	27
Bijlage A	Bestudeerde documentatie	29
Bijlage B	Geïnterviewde personen	33

Managementsamenvatting

Afbakening

Het duurzaam beheer van digitale data omvat verschillende begrippen. Onder 'data' wordt verstaan gestructureerde gegevens, in tegenstelling tot ongestructureerde gegevens ('bestanden', in de praktijk vaak documenten). Bij het 'duurzaam beheer' hiervan gaat het om de toegankelijkheid van de data vanaf ontstaan tot aan vernietiging of doorlopende preserving. Databeheer bevindt zich daarmee op het snijvlak tussen informatiebeheer (vaak gericht op bestandsbeheer) en gegevensmanagement (nog minder gericht op duurzame toegankelijkheid voor de langere termijn), twee disciplines die steeds verder in elkaar zullen (moeten) schuiven de komende jaren. Voor duurzaam databeheer bestaan verschillende juridische, ethische, organisatorische en technische kaders.

Juridische en ethische kaders

Zowel de Wob, de AVG en de Who als de Archiefwet stellen eisen over hoe er met data wordt omgegaan. Data valt of kan vallen onder de juridische begrippen 'document', uit de Wet openbaarheid van bestuur en de Wet hergebruik overheidsinformatie, 'persoonsgegevens' zoals beschreven in de AVG en 'archiefbescheiden' binnen de Archiefwet. Deze generieke wettelijke kaders zijn dus, in principe, onverkort van toepassing op data, nog los van sectorale wet- en regelgeving – een constatering die wellicht voor de hand ligt, maar verstrekkende gevolgen kan hebben voor Rijksoverheidsorganisaties. Kijkend vanuit het perspectief van deze wetten is er in principe geen relevant verschil tussen data en bijvoorbeeld een papieren document. Het medium, de opslagplaats, of de gebruikte techniek van data is voor de wet weinig relevant. Dit betekent dat er voor data dezelfde eisen gelden als voor meer traditionele vormen van informatieproducten. Op grond van de Wob moeten Rijksoverheidsorganisaties openbaarheid geven over hun besluitvorming. Wanneer burgers om data vragen, moeten organisatie hieraan voldoen binnen de door de wet gestelde termijnen. De Who stelt de randvoorwaarden op voor het hergebruik van openbare data van Rijksoverheidsorganisaties en stelt ook eisen aan het formaat waarin deze informatie beschikbaar moet worden gesteld. Op grond van de AVG moeten Rijksoverheidsorganisaties op een behoorlijke manier met persoonsgegevens omgaan en hiervoor technische, organisatorische en juridische waarborgen realiseren. Op grond van de Archiefwet moeten Rijksoverheidsorganisaties data in goede, geordende en toegankelijke staat brengen en bewaren en hun data vernietigen of overbrengen wanneer dat nodig is op grond van de selectielijsten. De ethische kaders voor datagebruik zijn nog in ontwikkeling en richten zich vaak op bepaalde aspecten rondom data (bijv. algoritmen). Te denken valt aan de KI-richtlijnen van de Europese Commissie, DEDA en de Toolbox innovatie Rijksoverheid. Deze kaders bouwen voort op bestaande wetgeving. Zij richten zich nog niet specifiek op digitaal duurzaam databeheer, maar raken dit onderwerp wel.

Organisatorische aspecten

Om te voldoen aan de wettelijke kaders moeten Rijksorganisaties de duurzame toegankelijkheid van data gedurende de hele levenscyclus waarborgen, vanaf ontvangst of creatie tot verzameling, bewerking, gebruik en uiteindelijk vernietiging of preserving. Het vraagt het nodige van Rijksorganisaties om dit in hun organisatie in te bedden. Dit kan door in te zetten op adequaat datamanagement binnen de organisatie. Hiervoor biedt de internationale Data Management Body of Knowledge (DMBOK) algemeen geaccepteerde richtlijnen. Hiernaast biedt het recent ontwikkelde datagovernancemodel voor de Rijksoverheid een doorvertaling van DMBOK voor de Rijksorganisaties, die hun ook inzicht geeft in hun eigen stand van zaken met betrekking tot datamanagement door middel van een zelfevaluatie-instrument. De NORA en GEMMA referentie-architecturen bieden bovendien een

gedetailleerde uitwerking van de verschillende rollen en verantwoordelijkheden die in het kader van datamanagement moeten worden belegd in de organisatie. Wel lijkt het erop dat het onderwerp duurzame toegankelijkheid daarbij nog enige aandacht behoeft vooraan in de levenscyclus van data en dat de rol van de archivaris of databeheerder naast bijvoorbeeld de CISO en de FG nog onderbelicht is. Dat is explicieter uitgewerkt in de Baseline informatiehuishouding Rijksoverheid. Ook voor de rol van de reguliere medewerker bij databeheer bestaan inmiddels bewustwordingsinstrumenten.

Technische aspecten

Duurzaam databeheer kent weliswaar technische uitdagingen, zoals databasebeheer, tijdreizen en vernietiging, maar techniek is volgend ten aanzien van de juridische en organisatorische eisen. Pas als Rijksoverheidsorganisaties die eisen hebben doorvertaald naar hun eigen organisatie, met behulp van een selectielijst voor data, is immers duidelijk wat er technisch nodig is. Niet voor alle uitdagingen zijn oplossingen beschikbaar, maar voor een goede invulling van databeheer bieden de DuTo-eisen van het Nationaal Archief, selectielijsten, de bestaande metadatamodellen, de FAIR-principes, NORA en de NEN-ISO normen nuttige generieke handvatten voor Rijksorganisaties. Uitgangspunt voor de duurzame toegankelijkheid is 'archiving by design': systemen van de Rijksoverheid dienen aan de voorkant zo ingericht te worden dat data duurzaam toegankelijk is gedurende de hele levenscyclus. Dat betekent dat er ook conserverings- en vernietigingsfunctionaliteit voor data voorhanden zal moeten zijn. Specifieke technische uitdagingen die aandacht behoeven zijn bijvoorbeeld de duurzame toegankelijkheid van relationele databases (die onder meer het archiveren van relaties tussen diverse gegevensbestanden vereist) en van algoritmes (waarbij wellicht trainings- en testdata ook bewaard moet blijven). Ook voor dit soort uitdagingen zijn technische oplossingen te vinden of te bedenken, maar ook hiervoor geldt dat eerst de vraag moet worden beantwoord welke data nu daadwerkelijk moet worden beheerd, vóórdat technische oplossingen in beeld komen.

Casus: databeheer bij het CBS

Het CBS verwerkt op grond van een wettelijke taak data van overheidspartijen. Na binnenkomst wordt de ruwe data binnen verschillende afdelingen bewerkt, geanonimiseerd, gemetadateerd en gekoppeld. Het resultaat van deze bewerkingen is het echte 'goud' van het CBS: microdata van hoge kwaliteit voor statistieken en onderzoek. Dit goud wordt opgeslagen in de schatkamer van het Data Service Centrum, het centrale data-archief. Binnen dit proces zijn verschillende rollen gedefinieerd. Het afdelingshoofd is als proceseigenaar verantwoordelijk voor de kwaliteit van en toegang tot de data; speciale medewerkers anonimiseren de data, en op verschillende momenten vinden deels automatische controles plaats. Metadatering gebeurt bij iedere stap in het werkproces en is verplicht voor alle betrokken medewerkers. Op technisch vlak slaat het CBS op verschillende momenten databases, of de wijzigingen hierin, op in het centrale archief. Data wordt bovendien, met metadata, omgezet in een eenvoudig (CSV of ASCII) formaat voor conserveringsdoeleinden. De belangrijkste lessen van het CBS op het gebied van duurzaam databeheer hebben betrekking op de manier waarop dit beheer is ingebed in de organisatie:

- 1) hanteer een **selectielijst** als basis voor databeheer;
- 2) maak **afspraken over eigenaarschap** van de data om verantwoordelijkheid te creëren;
- 3) richt **centrale opslag** in t.b.v. de datakwaliteit en duurzame toegankelijkheid;
- 4) organiseer verplichte **metadatering** met toezicht en verantwoordelijkheid op hoog niveau
- 5) creëer een **cultuur** voor de juiste omgang met data.

1. Inleiding

1.1 Aanleiding

Het rapport 'Ongekend Onrecht' van de Parlementaire ondervragingscommissie Kinderopvangtoeslag maakt duidelijk dat de informatievoorziening van de overheid aan de Tweede Kamer en aan de burger valt of staat met een goede informatiehuishouding¹. In de reactie op dit rapport² stelt het demissionaire kabinet een aantal maatregelen voor op dit gebied. Onderdeel hiervan is een generiek actieplan ter verbetering van de informatiehuishouding bij het Rijk, dat wordt uitgewerkt in een aantal departementale actieplannen.

Het generieke actieplan *'stelt kaders voor de informatiehuishouding voor het hele Rijk. Er zal onder andere in staan hoe ambtelijke en bestuurlijke besluitvorming wordt vastgelegd. De basis voor het actieplan is een analyse van de huidige stand van de informatiehuishouding. [...] Het actieplan wordt gemaakt onder regie van het overleg van de Secretarissen-generaal en is op 1 april klaar. Het actieplan bevat concrete maatregelen voor de verbetering van de informatiehuishouding. Daarbij maken we onderscheid tussen de rijksbrede verantwoordelijkheid en de verantwoordelijkheid van de individuele ministeries. Het actieplan geeft tevens inzicht in de kosten.'*³

Het lopende Rijksprogramma voor Duurzaam Digitale Informatiehuishouding (RDDI) heeft de vraag gekregen dit generieke actieplan op te stellen. Daarin zal ook het onderwerp 'data' aan bod komen.

Data wordt de afgelopen jaren wel genoemd als 'het nieuwe goud' waarop private organisaties hun businessmodellen kunnen bouwen, maar waarmee ook voor publieke organisaties veel voordeel te behalen valt, bijvoorbeeld in de dienstverlening aan burgers of simpelweg in de meetbaarheid, efficiency en effectiviteit van beleid. De hoeveelheid gegenereerde data is wereldwijd dermate snel toegenomen dat er wel wordt gesproken van een 'gedataficeerde' samenleving. Bij de overheid moet data hierdoor beschikbaar zijn voor veel meer doeleinden dan vroeger. Daarbij valt niet alleen te denken aan verschillende initiatieven rondom *smart cities* en datagedreven werken (één van de prioriteiten voor de informatiehuishouding van het Rijk⁴), maar ook aan de ontwikkelingen rondom kunstmatige intelligentie en de inzet van algoritmes door overheden. Hierdoor wordt het adequate beheer van al die data steeds belangrijker. Het rapport over de informatiehuishouding van de toekomst, dat Reframing Studio recent heeft opgesteld voor het project Future Lab van RDDI, zegt hier het volgende over:

Data is [...] niet alleen een nieuwe manier om de wereld te leren kennen maar ook om haar te besturen. Naarmate data belangrijker wordt in het nemen van beslissingen neemt het belang van databeheer ook toe⁵.

Data moet dus niet alleen beschikbaar zijn voor de datagedreven bedrijfsvoering van de overheid, maar ook voor het afleggen van verantwoording over beleid en uitvoering. Gedegen databeheer is voorwaardelijk voor het gebruik van data. In dit kader is het dan ook van belang de kaders voor het beheer van data scherp in beeld te hebben.

¹ Verslag parlementaire ondervragingscommissie. Ongekend onrecht, 17 december 2020.

² Kabinetsreactie rapport 'Ongekend onrecht', 15 januari 2021.

³ Kabinetsreactie rapport 'Ongekend onrecht', 15 januari 2021, p. 22.

⁴ RDDI, *Samen op weg naar duurzaam toegankelijke overheidsinformatie. Meerjarenplan verbetering informatiehuishouding Rijksoverheid. Actualisering ten behoeve van 2020*, januari 2020, p. 12 (plateau 3).

⁵ Reframing studio en Future Lab, *De informatiehuishouding van het Rijk in 2030*, april 2020, p. 36.

Vanuit RDDI heeft PBLQ de vraag gekregen om, op basis van een korte verkenning naar het onderwerp 'databeheer', deze kaders te inventariseren en te beschrijven en daarmee input aan te leveren voor het onderdeel 'data' uit het generieke actieplan, met een focus op de duurzame toegankelijkheid van data.

1.2 Opdrachtformulering en scope

Naar aanleiding van het bovenstaande is de opdracht aan PBLQ als volgt geformuleerd:

Het inventariseren en beschrijven van de kaders en uitgangspunten voor databeheer bij Rijkspartijen. Het gaat daarbij om juridische kaders, maar mogelijk ook om andersoortige uitgangspunten, zoals kwaliteitsstandaarden of technische, organisatorische en mogelijk ethische vereisten hiervoor. De verkenning moet antwoord geven op de vraag aan welke eisen het beheer van data bij het Rijk moet voldoen om deze data duurzaam toegankelijk te maken. De resultaten dienen te worden opgeleverd in de vorm van een rapport en door vertaald in een twopager.

In de verkenning dient ook het onderwerp algoritmes in de context van de Archiefwet meegenomen te worden. Daarnaast dient de verkenning een casus op het gebied van duurzaam digitaal databeheer te beschrijven. De verkenning dient bovendien op korte termijn opgeleverd te worden, om te kunnen dienen als input voor het boven genoemde generieke actieplan.

Bovenstaande betekent dat deze verkenning geen diepte-analyse biedt met betrekking tot duurzame toegankelijkheid van data, maar een overzicht op hoofdlijnen geeft van de *bestaande* kaders en normen voor de duurzame toegankelijkheid van data en hiernaar zal verwijzen. Gezien de beperkte duur en diepgang van het onderzoek bevat het rapport daarom ook geen aanbevelingen; wel worden her en der aandachtspunten benoemd.

Dat betekent eveneens dat allerlei ontwikkelingen die zich momenteel voordoen op het gebied van data en duurzame toegankelijkheid, maar die (nog) geen weerslag hebben op deze normen en kaders – zoals de ontwikkelingen rondom Common Ground, de veranderende inzichten binnen het vakgebied van het gegevensmanagement, de nieuwe Archiefwet en de Wet Open Overheid of de toenemende versmelting van data en bestanden – in deze verkenning grotendeels buiten beschouwing zullen blijven. Voor een dergelijke, meer toekomst-georiënteerde verkenning is nader onderzoek nodig. Het eerder aangehaalde rapport van Futurelab biedt hiervoor overigens wel aanzetten.

De focus van deze verkenning ligt tenslotte (zoveel mogelijk) op het beheer van data in het primair proces (primaire data), niet op het hergebruik van data binnen een nieuwe context, bijvoorbeeld in het kader van datagedreven werken (secundaire data) of beleidsontwikkeling.

1.3 Werkwijze

Deze verkenning heeft plaats gevonden in week 6 tot en met 8 van 2021 en bestond uit drie fasen:

Fase 1: verkennen en afbakenen

Deze fase bestond uit een kick-off en een eerste afbakening van het onderwerp databeheer. Hiervoor is deskresearch uitgevoerd. Bijlage A geeft een overzicht van de gebruikte achtergronddocumentatie.

Fase 2: verdiepen en analyseren

In deze fase hebben we de belangrijkste aspecten van databeheer nader onderzocht en uitgediept in enkele gesprekken met experts van ICTU, de Inspectie Overheidsinformatie en Erfgoed en het Nationaal Archief. Ook is hierbij de casus van het CBS bestudeerd aan de hand van een diepte-interview met enkele vertegenwoordigers van deze organisatie. Bijlage B geeft een overzicht van de geïnterviewde personen.

Fase 3: verwerken en rapporteren

In deze afsluitende fase is de opgehaalde informatie verwerkt, verrijkt en geduid om te komen tot deze eindrapportage en tot de twopager als afgeleide hiervan. Daarbij is de twopager voorgelegd aan de geïnterviewden ter validatie.

Iedere fase heeft grofweg één week beslagen. In week 6 (kick-off) en week 8 (voortgang) heeft afstemming plaats gevonden met de opdrachtgever.

1.4 Leeswijzer

Hieronder zetten wij eerst uiteen wat wij in het kader van deze beknopte verkenning verstaan onder data en onder het duurzaam beheer hiervan (hoofdstuk 2). Vervolgens behandelen wij de verschillende normen en kaders die daarbij een rol spelen, op juridisch en ethisch vlak (hoofdstuk 3), op organisatorisch vlak (hoofdstuk 4) en op technisch gebied (hoofdstuk 5). Per hoofdstuk worden na een korte inleiding de uitdagingen en de bestaande normen en kaders besproken, waarna ieder hoofdstuk wordt afgesloten met beknopte samenvatting. In het afsluitende hoofdstuk 6 beschrijven we de manier waarop het CBS als grote dataverwerkende organisatie de duurzame toegankelijkheid van data heeft georganiseerd en welke lessen hieruit te trekken zijn voor Rijksoverheidsorganisaties. In de bijlage geven we een overzicht van de achtergronddocumenten die wij hebben bestudeerd en van personen die wij hebben gesproken in het kader van deze verkenning. De managementsamenvatting van dit rapport wordt ook separaat opgeleverd als twopager.

2. Definities en afbakening

2.1 Inleiding

Hoewel het begrip 'data' veel wordt gebruikt in relatie tot de digitale overheid, is op voorhand niet altijd duidelijk wat met deze term wordt bedoeld. Nog los van het feit dat er ook wordt gesproken over 'gegevens' in het Nederlands, lijken de verschillen tussen met name data en informatie soms onduidelijk en worden beide termen door elkaar gebruikt. Ook de term 'beheer' kan, al naar gelang de context, soms verschillende betekenissen hebben. Zo kan de term betrekking hebben op het beheer van informatie voor de zeer lange termijn, en dan synoniem zijn van de term 'archivering', maar bijvoorbeeld ook slaan op beheeracties die binnen de bedrijfsvoering worden ondernomen bij het werken met data, bijvoorbeeld ter verbetering van de datakwaliteit. Iets dergelijks geldt ook voor de termen 'duurzaam' en 'duurzame toegankelijkheid'. Dit hoofdstuk gaat daarom kort in op de genoemde termen. Op die manier wordt niet alleen helder wat in het kader van deze verkenning onder deze begrippen wordt verstaan, maar wordt ook de verkenning als zodanig afgebakend.

2.2 Definities

Data

Er kan onderscheid worden gemaakt, zoals traditioneel in het informatiemanagement, tussen 'data', 'informatie' en 'kennis': losse data krijgt (gecombineerd) betekenis in de vorm van informatie, die kan leiden tot kennis⁶. De term 'data' is daarbij synoniem voor 'gegevens'⁷. Een voorbeeld hiervan is een tabel met lengtes en leeftijden van een populatie jongeren. De data is hierbij de leeftijd en lengte van de jongeren; wordt deze data bewerkt tot gemiddelden en vergeleken, dan ontstaat informatie, en deze informatie kan leiden tot kennis, namelijk tot het inzicht dat jongeren naarmate ze ouder worden veelal ook langer zijn.

Op grond hiervan hanteren we de definitie van 'data' zoals vastgesteld in de Archiefterminologie voor Nederland en Vlaanderen en gebruikt in de NORA⁸:

Weergave van een feit, begrip of aanwijzing, geschikt voor overdracht, interpretatie of verwerking door een persoon of apparaat.

Data is vaak gestructureerd in een database, informatie daarentegen is 'ongestructureerd', want daarbij gaat het om het gebruik van data in een bredere context dan alleen een database. Informatie heeft de vorm van een 'bestand' (traditioneel een document).

Beheer

Het beheer van data beslaat in het kader van deze verkenning de hele levenscyclus van data, van de creatie, opslag, verrijking en gebruik tot en met de vernietiging (of het eventuele hergebruik) of de duurzame bewaring voor de lange termijn of zelfs de eeuwigheid, en alle acties die daarbij horen. Daarmee valt het begrip in grote lijnen samen met het begrip 'datamanagement' of

⁶ https://en.wikipedia.org/wiki/DIKW_pyramid

⁷ We gebruiken beide termen door elkaar. De Nederlandse term het voordeel heeft dat het ook in het enkelvoud gebruikt kan worden, hoewel het langzaam gebruikelijk is geworden ook 'data' als een enkelvoud te beschouwen ('de data is...'), en wij dat in deze verkenning ook doen.

⁸ <https://www.noraonline.nl/wiki/Gegeven>

gegevensmanagement⁹. ‘Beheer’ is dus niet alleen de ‘archivering’ van data in de laatste levensfase, maar begint al op het moment van ontstaan van data.

Duurzame toegankelijkheid

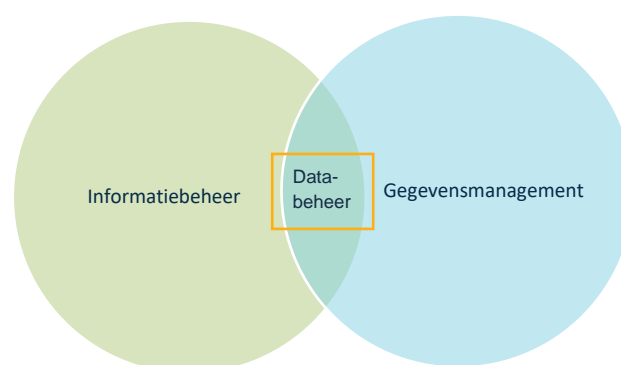
Voor dit begrip hanteren wij de definitie van het Nationaal Archief:

Toegankelijk betekent vindbaar, beschikbaar, leesbaar, interpreteerbaar en betrouwbaar voor degenen die er recht op hebben, vanaf het moment van ontstaan en voor zolang als noodzakelijk. Duurzaam betekent dat de toegankelijkheid van de informatie bestand is tegen veranderingen van elke aard¹⁰.

Duurzame toegankelijkheid gaat dus niet alleen over beheer op de lange termijn (na eventuele overbrenging in de zin van de Archiefwet), maar ook op de kortere termijn (in de eerste twintig – en onder de nieuwe Archiefwet tien – jaar na ontstaan). Het betekent, met andere woorden, dat ‘overheidsinformatie, gedurende de termijn waarin deze bewaard moet worden, in goede, geordende en toegankelijk staat moet blijven’.¹¹

2.3 Afbakening van deze verkenning

Gelet op het voorgaande heeft deze verkenning dus betrekking op het beheer van gestructureerde gegevens, in het bijzonder waar het beheer van deze gegevens afwijkt van het beheer van informatie in de vorm van bestanden. De focus ligt daarbij op de duurzame toegankelijkheid van data en op hetgeen moet worden georganiseerd om die te realiseren. Deze verkenning bevindt zich daarmee op het snijvlak tussen informatiebeheer (dat nog vaak gericht is op bestands- of documentbeheer) en gegevensmanagement (waarbinnen duurzame toegankelijkheid voor met name de langere termijn nog niet helemaal ontwikkeld lijkt) – twee disciplines die de komende tijd steeds meer zullen (moeten) integreren:



Dit rapport beschrijft de ‘status quo’: de grenzen tussen data en bestanden zijn aan het vervagen en dit stelt mogelijk nieuwe eisen aan het denken over ‘archief’, maar het is niet de opzet deze bredere – en

⁹ Zie hiervoor hoofdstuk 4. Overigens is ‘databeheer’ ook de Nederlandse vertaling van de Engelse term ‘datamanagement’.

¹⁰ <https://www.nationaalarchief.nl/archiveren/kennisbank/overzicht-van-begrippen>

¹¹ Inspectie Overheidsinformatie en Erfgoed, *Landelijke voorzieningen en WOZ en Adressen en Gebouwen*, april 2020, p. 16.

op zichzelf zeker gerechtvaardigde – vraag naar nut en noodzaak van het onderscheid tussen data en informatie in de vorm van bestanden hier te behandelen. Het gaat er primair om een beeld te geven van de bestaande kaders. Zoals hierboven in hoofdstuk 1 aangegeven, ligt de focus bovendien (zoveel mogelijk) op het beheer van data in het primair proces (primaire data), niet op hergebruikte data (secundaire data).

We benaderen het aldus afgebakende onderwerp hieronder vanuit drie perspectieven, van breed en abstract naar smal en concreet:

- een juridisch (en ethisch) perspectief (waar moet een Rijksorganisatie aan *voldoen?*) (hoofdstuk 3);
- een organisatorisch perspectief (wat moet een Rijksorganisatie daarvoor in de eigen organisatie *doen?*) (hoofdstuk 4);
- en een technisch perspectief (wat moet een Rijksorganisatie daarvoor op technisch gebied *doen?*) (hoofdstuk 5).

We vertalen de juridische kaders dus door naar organisatorische consequenties en uiteindelijk, voor zover mogelijk, naar technische randvoorwaarden. Daarbij kijken we per perspectief telkens naar de uitdagingen die hierbij gelden en naar de bestaande normen kaders, waarna we een korte samenvatting geven van de bevindingen.

3. Juridische en ethische kaders voor duurzame toegankelijkheid van data

3.1 Inleiding

De belangrijkste soort normen die gelden voor duurzaam databeheer zijn juridisch van aard. In dit hoofdstuk schetsen we daarom, op hoofdlijnen, de juridische kaders voor databeheer die zijn gesteld in de Wet openbaar bestuur (Wob), de Algemene Verordening Gegevensbescherming (AVG) en de Wet hergebruik overheidsinformatie. Daarbij gaan we niet alleen in op de belangrijkste wet- en regelgeving in dit verband, maar besteden we ook aandacht aan de ethische aspecten van duurzaam databeheer. Eisen uit andere mogelijke relevante (informatie) wetten, zoals de auteurswet, databankenwet, en materiewetten die door de aard van bepaalde data relevant kunnen zijn, worden buiten beschouwing gelaten, net als sectorale wet- en regelgeving.

3.2 Uitdagingen

Al in 1991 constateerde de Algemene Rekenkamer dat het beheer van gegevens onvoldoende aandacht had bij Rijksorganisaties, met alle risico's van dien¹². Gegevens werden simpelweg niet beschouwd als onderdeel van het archief, dat was gericht op (papieren) documenten. Anno 2021 is het duurzame beheer van data nog steeds een vraagstuk, terwijl data voor het handelen van de overheid steeds belangrijker wordt.

De hoofdvraag hier is wat vanuit juridisch oogpunt de status is van data en of data op één voet kan of moet worden gesteld met 'informatie' in de meer traditionele vorm van (digitale) documenten. Als de juridische status van data duidelijk is, wordt pas duidelijk hoe met data omgegaan dient te worden. Een tweede, hieraan verwante vraag heeft betrek op de meer ethische overwegingen rondom data. Wellicht *wil of zou* een organisatie immers niet alle data (*moeten*) beheren die zij wettelijk gezien *mag* beheren.

3.3 Normen en kaders

De belangrijkste *generieke* wettelijke kaders voor overheidsinformatie zijn de Archiefwet, de Wet openbaarheid van bestuur, de AVG en de Wet hergebruik overheidsinformatie¹³. Daarnaast is er veel *sectorale* regelgeving met impact op het beheer van overheidsinformatie, zoals de gemeentewet, de auteurswet en de aanstaande omgevingswet. Voor een overzicht van de belangrijkste wet- en regelgeving met betrekking tot archivering heeft KVAN-Brain in samenwerking met een aantal andere partijen de 'Wettenkaart' opgesteld.¹⁴ Op de site van KIA is een uitgebreider overzicht te vinden¹⁵. Ook de NORA bevat een overzicht van wet- en regelgeving die van invloed is op gegevensmanagement¹⁶. Hieronder gaan wij na wat deze belangrijkste generieke wettelijke kaders zeggen over data.

¹² Algemene Rekenkamer, *Machineleesbare gegevensbestanden; archivering en beheer bij het Rijk*, 1991.

¹³ We laten de Baseline Informatiebeveiliging Overheid hier buiten beschouwing.

¹⁴ <https://kia.pleio.nl/blog/view/55812212/de-wettenkaart-online>

¹⁵ <https://kia.pleio.nl/files/view/55811587/overzicht-archiefregelingen-peter-brand.xlsx>

¹⁶ https://www.noraonline.nl/wiki/Wettelijke_kaders_en_regelgeving_gegevensmanagement

3.3.1 Wet openbaarheid van bestuur

Volgens de grondwet moet de overheid bij *de uitvoering van haar taak openbaarheid volgens de regels bij de wet te stellen*¹⁷. De Wet openbaarheid van bestuur (Wob) regelt de wijze van openbaarmaking van informatie van bestuursorganen. Het stelt burgers in staat informatie op te vragen over het handelen van de overheid. In de wet wordt verwezen naar documenten. Hiermee wordt bedoeld: *'een bij een bestuursorgaan berustend schriftelijk stuk of ander materiaal dat gegevens bevat'*¹⁸.

Onder deze definitie vallen data/gegevens zoals gedefinieerd in hoofdstuk 2 van dit rapport. Recent is nog eens bevestigd dat hier behalve schriftelijke stukken ook, geluidsbanden, videobanden, ponskaarten, diskettes, cd-rom's, fotomateriaal, e-mailberichten, digitale informatie én sms- WhatsApp-berichten kunnen vallen¹⁹. Deze opsomming is niet volledig, want er kunnen nieuwe vormen van documenten ontstaan, door bijvoorbeeld nieuwe computertechnieken. De plek waarop iets is opgeslagen, maakt ook niet uit: *'de techniek van opslaan is niet bepalend is voor de vraag of de Wob wel of niet van toepassing is. Of een digitaal document nu opgeslagen is op een harde schijf, een eigen dan wel externe server of in de cloud, het gaat nog steeds om documenten "van" het bestuursorgaan.'*²⁰

In principe is overheidsinformatie openbaar. Bestuursorganen verschaffen uit eigen beweging informatie over het beleid, inclusief de voorbereiding en de uitvoering, wanneer dit in het belang is van een goede en democratische bestuursvoering²¹. De informatie moet onder andere actueel en nauwkeurig zijn. Daarnaast kan een burger een verzoek om informatie over een bestuurlijke aangelegenheid opvragen²². Een bestuursorgaan hoeft in een aantal gevallen geen informatie te verstrekken, bijvoorbeeld wanneer dit de veiligheid zou kunnen schaden, het gaat om persoonlijke beleidsopvattingen of als de informatie bijzondere persoonsgegevens bevat²³.

Documenten, inclusief data, vallen onder de verantwoordelijkheid van het bestuursorgaan en de Wob zolang zij bij dat orgaan berusten, oftewel tot het moment van overbrenging. De overbrengingstermijn is (nu nog) 20 jaar.

NB. De Wob wordt op termijn vervangen door de Wet open overheid (WOO). Deze wet gaat overheidsorganisaties verplichten om een aantal categorieën informatie – waaronder wet- en regelgeving, bestuursstukken, organisatiegegevens, raadsinformatie, jaarplannen, onderzoeken en klachten – actief en heel vroeg openbaar te maken en dient daarmee ook ter verbetering van de informatiehuishouding.

3.3.2 Wet hergebruik van overheidsinformatie

De richtlijn 2003/98/EG inzake hergebruik van overheidsinformatie²⁴ was eerder in de Wob geïmplementeerd. Deze richtlijn is door een wijziging²⁵ aangepast. Naar aanleiding hiervan is er een

¹⁷ Art. 110 Grondwet.

¹⁸ Art. 1a Wob.

¹⁹ Memorie van Toelichting bij Wob en ECLI NL RvS 2019 899.

²⁰ ECLI NL RvS 2019 899.

²¹ Art. 8 Wob.

²² Art. 3 Wob.

²³ Zie art. 10 en 11 Wob.

²⁴ Richtlijn 2003/98/EG van het Europees parlement en de Raad van de Europese Unie van 17 november 2003 inzake hergebruik van overheidsinformatie.

²⁵ Richtlijn 2013/37/EU van het Europees parlement en de Raad.

specifieke wet opgesteld ten aanzien van hergebruik van overheidsinformatie, omdat de richtlijn zich richt op informatie die algemeen openbaar is, terwijl de Wob zich (alleen) richt op bestuursinformatie²⁶. De Wet hergebruik van overheidsinformatie (Who) geeft burgers de mogelijkheid een verzoek in te dienen bij publieke instellingen²⁷ om (al openbaar gemaakte) informatie te mogen hergebruiken. De Who stelt de randvoorwaarden op voor het hergebruik van openbare overheidsinformatie. Hierbij wordt uitgegaan van beschikbaarstelling naar aanleiding van een verzoek. In de Who wordt er gesproken van document. 'Document' wordt hier anders gedefinieerd dan in de Wob. De definitie in de Who is *een bij een met een publieke taak belaste instellingen berustend schriftelijk stuk of ander materiaal dat gegevens bevat*²⁸.

Instellingen moeten voor hergebruik beschikbare informatie verstrekken *zoals de informatie bij de met een publieke taak belaste instelling aanwezig is en voor zover mogelijk langs elektronische weg, in een open en machinaal leesbaar formaat, samen met de metadata, waarbij het formaat en de metadata voor zover mogelijk voldoen aan formele open standaarden*²⁹. Van documenten is sprake bij *eender welke inhoud, eender welk deel van een dergelijke inhoud en ongeacht het medium*³⁰. Digitale documenten kunnen ook digitaal opgeslagen datasets of gegevens bevatten. Bij de Who gaat het ook om documenten die toegankelijk zijn op grond van de Archiefwet, of op grond van regelgeving inzake openbare registers, of meteorologische data, of verkeersgegevens. Ook hier is de Who dus van toepassing op data.

Een document wordt als document in machineleesbaar formaat beschouwd als het een bestandsformaat heeft met een zodanige structuur dat softwaretoepassingen eenvoudig gegevens in het document kunnen identificeren, herkennen en extraheren³¹. Dit vraagt dus van organisatie om de informatie op zo'n manier te organiseren dat zij aan hergebruiksverzoeken kunnen voldoen.

De Who is onder andere van toepassing op informatie opgeslagen bij Rijksorganisaties én bij archiefinstellingen. Uit de artikelen 15, 16 en 17 van de Archiefwet blijkt welke archiefbescheiden die in een archiefbewaarplaats berusten openbaar zijn, en waar dus een hergebruikverzoek voor kan worden ingediend. Iedereen, inclusief bedrijven, is bevoegd die archiefbescheiden kosteloos te raadplegen en daarvan of daaruit afbeeldingen, afschriften, uittreksels en bewerkingen te maken of op zijn kosten te doen maken.

Er is een handleiding opgesteld voor publieke instellingen die informatie geeft over hoe zij invulling kunnen geven aan de Wet hergebruik van overheidsinformatie³².

3.3.3 AVG

De Algemene Verordening Gegevensbescherming (AVG) ziet toe op de bescherming van persoonsgegevens. De wet is dan ook alleen van toepassing op die data die ook aan te merken is als

²⁶ Memorie van toelichting bij de Wob.

²⁷ Een openbaar lichaam als bedoeld in art 1 lid 1 van de Richtlijn 2003/98/EG van het Europees parlement en de Raad van 17 november 2003 inzake het hergebruik van overheidsinformatie.

²⁸ Art. 1 sub d Who.

²⁹ Art. 5 Who.

³⁰ Art. 2 lid 3, Richtlijn 2003/98/EU van het Europees parlement en de raad.

³¹ Memorie van toelichting bij de Who.

³² https://open-overheid.nl/wp-content/uploads/2016/05/WEB_90943_BZK_Handleiding-Who-versie2.pdf

een persoonsgegeven. In de wet³³ wordt een persoonsgegeven als volgt gedefinieerd: *alle informatie over een geïdentificeerde of identificeerbare natuurlijke persoon*. Hierbij moet informatie breed worden geïnterpreteerd en valt het in hoofdstuk 1 gedefinieerde data dus óók onder het begrip informatie zoals beschreven in de AVG. Dit komt doordat het al snel gaat om persoonsgegevens. Gegevens die direct of indirect naar een persoon kunnen worden herleid zijn namelijk ook persoonsgegevens. Hierbij wordt in de AVG nadrukkelijk aangegeven dat dit kan door verwachte gegevens, zoals een naam en een identificatienummer, maar ook door gegevens die nog niet voor iedereen altijd herkenbaar zijn als identificeerbare gegevens, zoals locatiegegevens of hartslag.

De AVG maakt onderscheid tussen (gewone) persoonsgegevens en bijzondere persoonsgegevens. Bijzondere persoonsgegevens zijn persoonsgegevens waaruit ras of etnische afkomst, politieke opvattingen, religieuze of levensbeschouwelijke overtuigingen, of het lidmaatschap van een vakbond blijken, en verwerking van genetische gegevens, biometrische gegevens met het oog op de unieke identificatie van een persoon, of gegevens over gezondheid, of over iemands seksueel gedrag of gerichtheid. Verwerkingen van bijzondere persoonsgegevens zijn in principe verboden tenzij er aan een aantal voorwaarden is voldaan³⁴.

Om te bepalen wanneer gegevens identificatie mogelijk maken, moet worden gekeken naar onder andere de kosten van en de tijd die nodig is om aan de hand van gegevens iemand te identificeren. Ook moet worden meegenomen of er door middel van beschikbare (of binnenkort beschikbare) technologie alsnog kan leiden tot identificatie³⁵. Over het algemeen geldt dat hoe meer data er is in relatie tot een persoon, hoe makkelijker het is om met deze data een persoon te identificeren. Dit is ook mogelijk wanneer grote hoeveelheden geanonimiseerde of gepseudonimiseerde data wordt gecombineerd. Daarnaast blijkt steeds vaker dat initieel onpersoonlijke gegevens, zoals hartslag of stemherkenning, ook tot identificatie kunnen leiden. Hierdoor is er bij grote hoeveelheden data snel sprake van persoonsgegevens. Op anonieme data die niet kan leiden tot identificatie van een persoon is de AVG niet van toepassing.

Wanneer data persoonsgegevens zijn, betekent dit voor overheidsinstantie dat zij een wettelijke grondslag nodig hebben om deze data te kunnen verwerken³⁶. Er moet een verwerkingsregister worden opgesteld waar duidelijk in staat welke verwerkingen plaatsvinden, op basis van welke grondslag, met welk doel en hoe lang de gegevens worden bewaard. Daarnaast stelt de AVG eisen aan het ontwerp en gebruik van informatiesystemen en -processen met betrekking tot persoonsgegevens. Deze eisen overlappen deels met opgetelde normenkaders zoals de Baseline Informatiebeveiliging Overheid (BIO). De BIO is breder van toepassing, aangezien die meer omvat dan alleen *persoonsgegevens*. Ook organisatorisch brengt de AVG een verplichting met zich mee, aangezien iedere organisatie moet beschikken over een Functionaris Gegevensbescherming (FG).

De AVG geeft burgers ook een aantal rechten ten aanzien van hun persoonsgegevens. In bepaalde omstandigheden kunnen zij hun gegevens opvragen, vragen om aanpassingen of vragen om hun gegevens te laten verwijderen. Het overtreden van de AVG kan leiden tot hoge boetes.

Er zijn verschillende handreikingen en handleidingen die Rijksoverheidsorganisatie helpen te voldoen aan de AVG (en verwante wetgeving). De Handleiding Algemene Verordening gegevensbescherming

³³ Art. 4 lid 1 AVG.

³⁴ Art. 9 AVG.

³⁵ Overweging 26 AVG.

³⁶ Art. 6 AVG.

legt uit wat de AVG betekent voor partijen die persoonsgegevens verwerken en aan welke regels de gegevens verwerking moet voldoen³⁷. Ook is de Rijksoverheid is verplicht om bij de ontwikkeling van nieuwe wetgeving rekening te houden met de resultaten van een Privacy Impact Assessment (PIA). Een Data protection impact assessment (DPIA, ook wel PIA of gegevensbeschermingseffectbeoordeling) is een instrument om vooraf de privacyrisico's van gegevensverwerking in kaart te brengen en maatregelen te kunnen nemen om deze risico's te verkleinen. Dit is een onderzoek naar de effecten van de wetgeving op de gegevensbescherming van burgers. Om organisaties hierbij te helpen is een Model gegevensbeschermingseffectbeoordeling Rijksdienst opgesteld³⁸.

3.3.4 Archiefwet

De Archiefwet heeft als doel de bevordering van efficiënte bedrijfsvoering en geheugen, de borging van de authenticiteit en integriteit van (overheids)informatie, de bevordering van transparantie van de overheid en het afleggen van verantwoording en mogelijk maken van de reconstructie van het verleden en het veiligstellen van cultureel erfgoed.

De wet is van toepassing op archiefbescheiden bij overheidsorganen. Dat zijn volgens de wet: *bescheiden, ongeacht hun vorm, door de overheidsorganen ontvangen of opgemaakt en naar hun aard bestemd daaronder te berusten*³⁹. Met de toevoeging 'ongeacht de vorm' heeft de wetgever beoogd om alle mogelijke soorten bescheiden onder het regime van de Archiefwet te scharen. Archiefbescheiden lijkt op het begrip 'document' uit de Wob, maar is nóg breder. Hieronder valt dus ook data zoals gedefinieerd in hoofdstuk 1.

Overheidsorganen zijn bovendien *verplicht de onder hen berustende archiefbescheiden in goede, geordende en toegankelijke staat te brengen en te bewaren, alsmede zorg te dragen voor de vernietiging van de daarvoor in aanmerking komende archiefbescheiden*⁴⁰. Dit betekent dat ook wanneer data nog valt onder het Wob-regime, overheidsorganisaties al verplicht zijn deze in een goede, geordende en toegankelijke staat te brengen.

In selectielijsten wordt bepaald welke archiefbescheiden moeten worden bewaard en vernietigd⁴¹. Archiefbescheiden die voor permanente bewaring in aanmerking komen op grond van de selectielijst, dienen te worden 'overgebracht' naar een archiefbewaarplaats na 20 jaar. Het Archiefbesluit en de Archiefregeling geven nadere regels voor het opstellen van selectielijsten en de duurzaamheid en toegankelijkheid van digitale archiefbescheiden. In een selectielijst wordt duidelijk om welke categorieën informatie het gaat en welke bewaartermijnen (in de zin van de AVG) worden gehanteerd⁴². De uitgangspunten van de AVG zijn nog niet altijd al meegenomen bij het opstellen van de selectielijst. In dat geval moeten zijn (nog) worden herzien.

Archiefbescheiden zijn na overbrenging openbaar, tenzij de openbaarheid conflicteert met andere belangen, zoals de eerbiediging van de persoonlijke levenssfeer, het belang van de Staat of het bedrijfsgeheim.

³⁷ <https://www.rijksoverheid.nl/documenten/rapporten/2018/01/22/handleiding-algemene-verordening-gegevensbescherming>

³⁸ <https://www.rijksoverheid.nl/documenten/rapporten/2017/09/29/model-gegevensbeschermingseffectbeoordeling-rijksdienst-pia>

³⁹ Art. 1c Archiefwet.

⁴⁰ Art. 3 Archiefwet.

⁴¹ Art. 5 Archiefwet.

⁴² KVAN/BRAIN, *Weten of vergeten?*, april 2020.

NB. De Archiefwet 1995 wordt op termijn vervangen door een nieuwe Archiefwet. Deze hanteert de term 'document' in plaats van 'archiefbescheiden' en beoogt met name een verkorting van de huidige overbrengstermijn van twintig jaar naar tien jaar, waardoor informatie eerder openbaar toegankelijk wordt. Ook worden de mogelijkheden voor 'bewaren bij de bron' uitgebreid door middel van een ontheffing van de overbrengingsverplichting. Dit wetsvoorstel dient daarmee ook tot verbetering van de informatiehuishouding van overheden.

3.3.5 Ethische overwegingen

Overheidsorganisaties mogen niet alle data die zij beheren ook daadwerkelijk gebruiken, of willen dit mogelijk zelf niet. Zij kunnen er bijvoorbeeld voor kiezen om geen tweede nationaliteit van burgers te registreren als dit geen toegevoegde waarde heeft binnen het primair proces of dit kan leiden tot ongewenst (her)gebruik en/of stigmatisering. Dat vereist ook iets voor de manier waarop data wordt opgeslagen en voor de toegang tot de data.

Er bestaan verschillende kaders, richtlijnen en normen die betrekking hebben op het gebruik van data door de overheid. Ethische overwegingen bij data laten zich lastig in een algemeen kader vangen, omdat 'data' over van alles kan gaan. Bij het werken met gezondheidsdata zullen immers andere overwegingen plaatsvinden dan als het gaat om geografische data. De bestaande normen en kaders zijn nog volop in ontwikkeling en focussen meestal op een of meer van de volgende aspecten: publieke waarden en grondrechten, de privacy van de burger, gelijke behandeling, het belang van transparantie, het belang van menselijke tussenkomst en organisatorische aspecten⁴³. Ze richten zich nog niet specifiek op de duurzame toegankelijkheid van data, maar raken hier wel aan.

Hieronder lichten we kort enkele van de belangrijkste (internationale) kaders binnen dit aandachtsgebied toe:

- Door de High-Level Expert Group on Artificial Intelligence is er een ethische richtlijn opgesteld voor AI⁴⁴. In dit richtsnoer wordt een kader voor het bewerkstelligen van betrouwbare KI geschetst. In het kader wordt niet expliciet ingegaan op de eerste component van betrouwbare KI, namelijk dat zij wettig moeten zijn door te voldoen aan alle toepasselijke wet- en regelgeving, maar in plaats daarvan is het bedoeld als richtsnoer voor het bevorderen en waarborgen van ethische en robuuste kunstmatige intelligentie.
- De Toolbox Ethisch Verantwoorde Innovatie⁴⁵ biedt ontwikkelaars en bestuurders bij de (Rijks)overheid inzicht in wat er nodig is voor innovatie met respect voor publieke waarden en grondrechten. Aan de hand van zeven heldere kernprincipes geeft de toolbox een aantal concrete adviezen, elk voorzien van verwijzingen naar 'tools' (hulpmiddelen, handreikingen, richtlijnen en voorbeelden).
- De Ethische Data Assistent (DEDA)⁴⁶ van de Universiteit Utrecht is een spelbord-game waarbij op een systematische wijze vragen moeten worden beantwoord over een project/programma op het gebied. Daarbij komen vraagstukken aan de orde op het gebied van security, informatiebeveiliging, integriteit en privacy.

⁴³ <https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/nieuwe-technologieen-data-en-ethiek/het-led/toolbox/aandacht-voor-ethische-en-juridische-aspecten/>

⁴⁴ *Ethische Richtsnoeren voor betrouwbare KI*, 8 april 2019 (<https://ec.europa.eu/futurium/en/ai-alliance-consultation>)

⁴⁵ <https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/nieuwe-technologieen-data-en-ethiek/publieke-waarden/toolbox-voor-ethisch-verantwoorde-innovatie/>

⁴⁶ <https://dataschool.nl/deda/>

3.4 Samenvatting

Zowel de Wob, de AVG en de Who als de Archiefwet stellen eisen over hoe er met data wordt omgegaan. Data valt of kan vallen onder de juridische begrippen ‘document’, uit de wet openbaarheid van bestuur, ‘persoonsgegevens’ zoals beschreven in de AVG en ‘archiefbescheiden’ afkomstig uit de Archiefwet. Kijkend vanuit het perspectief vanuit deze drie wetten is er in principe geen relevant verschil tussen data en bijvoorbeeld een papieren document. Het medium, de opslagplaats, of de gebruikte techniek van data is voor de wet weinig relevant.

Dit betekent dat voor data dezelfde eisen gelden als voor meer traditionele vormen van informatieproducten – een constatering die wellicht voor de hand ligt, maar verstrekkende gevolgen kan hebben voor Rijksorganisaties, die deze wettelijke kaders organisatorisch en technisch door moeten vertalen. Op grond van Wob moeten Rijksoverheidsorganisaties openbaarheid geven over hun besluitvorming. Wanneer burgers om informatie vragen, moeten organisatie hieraan voldoen binnen de door de wet gestelde termijnen. De Who stelt de randvoorwaarden op voor het hergebruik van openbare overheidsinformatie van (onder andere) Rijksoverheidsorganisaties. Hierbij wordt uitgegaan van beschikbaarstelling naar aanleiding van een verzoek. De Who stelt ook eisen hoe (in wat voor formaat) deze informatie beschikbaar moet worden gesteld.

Op grond van de AVG moeten Rijksoverheidsorganisatie op een behoorlijke manier met persoonsgegevens omgaan en hiervoor technische, organisatorische en juridische waarborgen realiseren. Op grond van de Archiefwet moeten Rijksoverheidsorganisaties data in goede, geordende en toegankelijke staat brengen en bewaren, ook voor overbrenging, en hun data vernietigen of overbrengen wanneer dat nodig is op grond van de selectielijsten.

De ethische kaders voor data zijn nog in ontwikkeling en richten zich vaak op bepaalde aspecten rondom data, zoals algoritmen. Deze kaders bouwen voort op bestaande wetgeving. Zij richten zich nog niet specifiek op digitaal duurzaam databeheer, maar raken dit onderwerp wel.

4. Organisatorische aspecten

4.1 Inleiding

Het adequaat beheren van data kan alleen wanneer de dataverwerkende organisatie hierop ingericht is. In dit hoofdstuk vertalen we daarom de juridische kaders door naar hetgeen Rijksoverheidsorganisaties moeten doen op organisatorisch vlak om de duurzame toegankelijkheid van hun data te waarborgen.

4.2 Uitdagingen

De grootste uitdaging op organisatorisch vlak is inzichtelijk te krijgen hoe welke data door de organisatie stroomt, van verzameling naar bewerking, gebruik en uiteindelijk vernietiging of permanente preservering⁴⁷. Die vraag wordt steeds prangender naarmate de beleidscyclus een datacyclus wordt⁴⁸ en besluiten vaker worden genomen op basis van data, waardoor de duurzame toegankelijkheid van data nodig is om verantwoording af te kunnen leggen. Daarbij komt een groot aantal vragen kijken: Welke data hebben we eigenlijk allemaal in huis? Waar en wanneer wordt data gecreëerd, gedeeld, bewerkt, verrijkt? Wie is bij deze handelingen betrokken en wie is verantwoordelijk? Welke handelingen moeten of mogen op welk moment worden uitgevoerd? Waar bevindt de data zich nadat deze handelingen zijn uitgevoerd? Wie heeft op welk moment toegang tot de data? En hoe blijft de data ook na langere tijd nog toegankelijk? Rijksorganisaties moeten deze vragen kunnen beantwoorden om aan de wettelijke kaders en normen te kunnen voldoen die zijn toegelicht in hoofdstuk 3.

4.3 Normen en kaders

Bovenstaande vragen gaan vooral over de manier waarop het werken met data is georganiseerd binnen Rijksoverheidsorganisaties. Hiervoor bestaan verschillende richtinggevende kaders. Ook is er recenter vanuit het Rijksprogramma Duurzaam Digitale Informatiehuishouding aandacht gekomen voor de rol die medewerkers zelf spelen bij het beheer van overheidsinformatie.

4.3.1 Datamanagement

Het organiseren van de manier waarop organisaties met data werken is het vakgebied van datamanagement. De Data Management Body Of Knowledge (DMBOK) is het internationaal erkende standaardwerk op het gebied van datamanagement, dat wordt uitgegeven door DAMA international. Dit referentiehandboek geeft de volgende definitie van datamanagement:

⁴⁷ Het Nationaal Archief definieert digitale preservering als volgt: *Het op zodanige wijze vastleggen, bewaren, beheren en beschikbaar stellen van digitale archiefbescheiden, dat deze ook na verloop van tijd raadpleegbaar, toegankelijk en authentiek zijn* (<https://www.nationaalarchief.nl/archiveren/kennisbank/wat-is-preservering>).

⁴⁸ <https://ibestuur.nl/podium/van-beleids-naar-datacyclus>

*Data management is the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets.*⁴⁹

Datamanagement omvat daarmee veel verschillende disciplines. DMBOK maakt onderscheid in de volgende deelgebieden binnen datamanagement:

- Data architectuur;
- Datamodellering en Ontwerp;
- Dataopslag en Beheer;
- Databeveiliging;
- Data-integratie en interoperabiliteit;
- Documenten en Content;
- Referentie- en Masterdata;
- Datawarehousing en BI;
- Metadata;
- Datakwaliteit.

Daarnaast benoemt DMBOK ook het thema datagovernance. In de benadering van DMBOK gaat het onderwerp datagovernance over sturing, planning, controle en toezicht op de omgang met data binnen een organisatie. Het is daardoor één van de deelgebieden van datamanagement, maar tegelijkertijd ook een doorsnijdend thema binnen het vakgebied, omdat de governance raakt aan alle andere deelgebieden van datamanagement. Om die reden is het geplaatst in het midden van het 'datamanagementwiel' van DAMA.

4.3.2 Datagovernance

Tegen deze achtergrond is binnen de Nederlandse Rijksoverheid onlangs een datagovernancemodel ontwikkeld in het kader van een Rijksbreed onderzoek naar datagovernance⁵⁰. Dit model, het 'Kenniswiel datagovernance', is 'ontwikkeld op basis van internationale standaarden, raakt aan negen kritieke functies en heeft de potentie om zich te ontwikkelen tot Rijksbrede standaard'⁵¹. De negen functies waarop het wiel ingaat zijn:

- Organisatievisie, -missie en -strategie;
- Legitimiteit: wet- en regelgeving;
- Sturing: organisatie en risicomangement;
- Processen, trainingen en medewerkers;
- Interactie met burgers en bedrijven;
- Data uitwisselingen met derden;

⁴⁹ DAMA-DMBOK Guide (Data Management Body of Knowledge). Datamanagement (of gegevensmanagement) is door NORA gedefinieerd als: *het integraal en beheerst verwerken van gegevens in een organisatie zowel op strategisch, tactisch als operationeel niveau met als doel de gewenste kwaliteit en beschikbaarheid te realiseren* ([https://www.noraonline.nl/wiki/Integraal_\(begrip_gegevensmanagement\)](https://www.noraonline.nl/wiki/Integraal_(begrip_gegevensmanagement))) en door GEMMA als 'het geheel van activiteiten om in de gemeente op het juiste moment over de juiste gegevens van de juiste kwaliteit te beschikken' (https://www.gemmaonline.nl/index.php/GMT_Gegevensmanagement).

⁵⁰ <https://led.pleio.nl/groups/view/fbdb70e5-2c2d-4a62-b6b1-55fe53e61e5c/datagedreven-werken-community/files>

⁵¹ <https://ibestuur.nl/podium/geen-woorden-maar-data>

- Dataverwerkingsmethodieken;
- Technologie voor dataverwerking;
- Datamanagement en datakwaliteit.

Per functie is ook een best practice bij de Rijksoverheid uitgelicht. Daarnaast is in het kader van dit onderzoek een instrument voor 'selfassessment' ontwikkeld⁵². Deze zogenaamde 'Positiebepaling Datagovernance' biedt inzicht in de stand van zaken op deze negen cruciale functies van datagovernance binnen de organisatie, in de ambitie en in de afstand tussen deze twee. Deze instrumenten bieden Rijksoverheidsorganisaties de mogelijkheid om met datagovernance aan de slag te gaan in hun eigen organisatie.

4.3.3 Gegevensmanagementrollen

De DMBOK en het Nederlandse datagovernancemodel bieden dus nuttige kaders voor de omgang met data binnen de organisatie. Daarnaast zijn binnen de expertgroep gegevensmanagement van NORA de rollen beschreven die een organisatie moet invullen om te komen tot adequaat gegevensmanagement⁵³. Hierbij wordt onderscheid gemaakt tussen de volgende rollen:

- Registratiehouder;
- Portefeuillehouder gegevensmanagement;
- Bronhouder;
- Regisseur gegevensmanagement;
- Distributeur;
- Afnemer;
- Architect;
- Gegevensmakelaar;
- Beheerder register/applicatie specifieke gegevensverzameling;
- Ontwikkelaar gegevensverzameling;
- Kwaliteitscontroleur;
- Functionaris voor de gegevensbescherming (FG).

In GEMMA zijn de rollen en verantwoordelijkheden verder uitgewerkt en wordt het gehele gegevensfunctiehuis beschreven in een RACI-matrix voor niet alleen gegevensmanagement, maar ook informatiebeveiliging, privacy en ENSIA-rollen⁵⁴. Opvallend is overigens dat de verantwoordelijkheid voor duurzame toegankelijkheid op dit moment niet expliciet belegd in deze rolbeschrijvingen, terwijl de archivaris vooraan in het proces van databeheer een plaats zou moeten hebben naast de CISO en de FG.

Duurzame toegankelijkheid is wel expliciet de insteek van de Baseline Informatiehuishouding Rijksoverheid⁵⁵. In dit normenkader, dat is gebaseerd op wet- en regelgeving en is bedoeld voor lijnmanagers, worden zeven normen gehanteerd en gekoppeld aan rollen binnen de organisatie.

⁵² <https://led.pleio.nl/groups/view/fbdb70e5-2c2d-4a62-b6b1-55fe53e61e5c/datagedreven-werken-community/files>

⁵³ https://www.noraonline.nl/wiki/Rollen_binnen_gegevensmanagement

⁵⁴ https://www.gemmaonline.nl/index.php/GM_Gegevensfunctiehuis

⁵⁵ <https://www.informatiehuishouding.nl/Producten+%26+publicaties/richtlijnen/2018/05/01/baseline-informatiehuishouding-rijksoverheid>

Daarmee is het ook een geschikt instrument om de kwaliteit van de informatiehuishouding een plaats te geven in de reguliere planning en controlcyclus van Rijksoverheidsorganisaties.

Om het strategische belang van data en datamanagement verder te concretiseren binnen Rijksoverheidsorganisaties wordt op dit moment overigens ook de rol van de CDO of Chief Data Officer ontwikkeld bij het Rijk⁵⁶.

4.3.4 De rol van de reguliere medewerker

Uiteraard heeft ook de 'gewone' medewerker een belangrijke rol te spelen bij databeheer en, breder, in de informatiehuishouding van Rijksoverheidsorganisaties. Wanneer zij op de juiste manier omgaan met data, vanuit het besef dat dit belangrijk is en een integraal onderdeel van hun werk, heeft direct positieve gevolgen voor de kwaliteit, vindbaarheid en bruikbaarheid van gegevens.

Vanuit het Rijksprogramma Duurzaam Digitale Informatiehuishouding heeft deze rol recent dan ook meer aandacht gekregen. Zo is er een 'kenniskaart' ontwikkeld voor Rijksambtenaren, die aangeeft hoe zij zorgvuldig kunnen werken met overheidsinformatie en deze informatie – in brede zin: traditionele documenten, maar ook gegevens in databanken en communicatie via e-mail of social media – ordelijk kunnen bewaren of vernietigen⁵⁷. Ook is er onlangs een bewustwordingscampagne rond dit thema gestart⁵⁸.

4.4 Samenvatting

Om data duurzaam te beheren volgens de wet- en regelgeving die hierop van toepassing is, is het nodig om de hele levenscyclus van data in beeld te hebben en de afzonderlijke momenten in die cyclus te koppelen aan rollen binnen de organisatie. Op die manier ontstaat duidelijkheid over de data die Rijksoverheidsorganisaties beheren én over verantwoordelijkheden, eigenaarschap van en toegang tot de data.

De bestaande kaders voor datamanagement en datagovernance (DMBOK, het datagovernancemodel voor de Rijksoverheid en NORA/GEMMA) vormen een nuttige basis om hierin inzicht te krijgen en om dit in te richten. Met betrekking tot duurzame toegankelijkheid is het van belang ook dit onderwerp expliciet te beleggen in de datamanagement of datagovernance strategie en om vooraan in de levenscyclus al na te denken over de toegankelijkheid van data op de langere termijn en hierbij de archivaris of een vergelijkbare rol in te schakelen naast bijvoorbeeld de CISO en de FG. De Baseline informatiehuishouding Rijksoverheid kan hierbij helpen. Ook bestaan er verschillende instrumenten die zich richten op de rol van reguliere medewerkers bij de Rijksoverheid.

De genoemde kaders zijn vanzelfsprekend geen garantie voor succes. Uiteindelijk staat of valt het beheer met de goede implementatie en uitvoering van deze kaders. Maar als Rijksoverheidsorganisaties aan de hand hiervan databeheer expliciet te beleggen binnen hun organisatie, hebben ze een goed uitgangspunt om duurzaam databeheer te realiseren.

⁵⁶ <https://led.pleio.nl/events/view/254fe0a0-bd1f-4ffe-8cce-92513598c4a4/chief-data-officer-een-waardevolle-positie-of-rol>

⁵⁷ <https://www.informatiehuishouding.nl/Producten+%26+publicaties/instrumenten/2019/09/17/kenniskaart-werken-met-overheidsinformatie>

⁵⁸ <https://www.informatiehuishouding.nl/voormedewerkers>

5. Technische aspecten

5.1 Inleiding

De wet- en regelgeving voor databeheer vertaalt zich niet alleen door in de inrichting van de organisatie, maar uiteindelijk ook in technische vereisten. De techniek is daarbij noodzakelijkerwijs *volgend*: Rijksorganisaties moeten in eerste instantie zelf bepalen welke data ze op welke manier willen of moeten beheren op grond van wet- en regelgeving om vervolgens te kunnen bepalen wat daar technisch voor nodig is. In dit hoofdstuk gaan we daarom vooral in op de generieke kaders die Rijksorganisaties kunnen gebruiken voor deze technische doorvertaling. Wel benoemen we enkele ‘blinde vlekken’ op technisch vlak die richting de toekomst nadere uitwerking behoeven.

5.2 Uitdagingen

Belangrijke uitdagingen op technisch vlak voor databeheer hebben allereerst betrekking op de kwaliteit van de data. Om de toegankelijkheid van data te garanderen in de hele levenscyclus moet data vindbaar en (her)bruikbaar zijn. Een absolute voorwaarde hiervoor is adequate metadatering: zonder metadata is data niet duurzaam toegankelijk. Toegang tot data is ook gekoppeld aan de opslagvorm (centraal, zoals bij de basisregistraties of kernregistraties van organisaties, decentraal of lokaal, zoals vakregistraties binnen afdelingen), waarbij het type opslag passend moet zijn bij het type gebruik van de data.

Bij regelmatig veranderende data is ook het *moment* van opslag belangrijk. Er kan voor worden gekozen rustpunten in de datastream in te brengen waarop de data wordt ‘bevroren’ en opgeslagen. Een andere mogelijkheid is om niet de volledige data telkens opnieuw op te slaan maar alleen de wijzigingen. Hiermee ontstaat een administratie van elke wijziging en kan de actuele dataset gereconstrueerd worden. Ongeacht de keuze moet het mogelijk zijn ‘in de tijd te reizen’: gebruikers moeten de mogelijkheid hebben om op een later moment terug te reizen naar de stand van zaken op een eerder moment. Dit maakt het mogelijk toestanden uit het verleden te reconstrueren en zo bijvoorbeeld ook verantwoording af te leggen, indien nodig. Daarbij kennen relationele databases extra uitdagingen. Een relationele database bevat diverse objecten die gezamenlijk via gedefinieerde relaties informatie bevatten. Voor duurzame toegankelijkheid is het nodig deze relaties te preserven, maar dat is complex, zodat informatieverlies op de loer ligt. Voor dit type database is het ook de vraag wat de eenheid van de informatie is, die gemetadateerd en opgeslagen moet worden (Ieder gegeven op zich? Een gehele rij? De hele database?), en wat de representatievorm is die moet worden bewaard (Databasecellen? Een export in XML? Of bijvoorbeeld een schermweergave?).

In dit kader zijn ook toegang en logging van groot belang. Dit zijn primair beveiligingsvraagstukken, die samenhangen met organisatorische rollen (wie heeft toegang tot welke data op grond van welke rol?), maar hieraan zit ook een technische component, omdat iedere wijziging later moet kunnen worden uitgelegd aan de hand van de loginformatie. Logging is dus een essentiële component voor de betrouwbaarheid van data en daarmee voor verantwoording.

Op grond van de AVG (en de Wob) is het van belang data te kunnen pseudonimiseren of anonimiseren. Ook dit is een technische uitdaging, zij het dat dit voor gestructureerde databases mogelijk eenvoudiger kan zijn dan voor ongestructureerde documenten, omdat in een gestructureerde database desbetreffende informatie gegroepeerd is. Een heel ander aandachtspunt is de vernietiging van data.

Om te kunnen voldoen aan de Archiefwet moet data onherstelbaar vernietigd kunnen worden. Systemen waarin data wordt opgeslagen, beschikken echter niet per definitie over de hiervoor benodigde functionaliteit⁵⁹. In het kader van de Archiefwet is ook preservering een belangrijk onderwerp. Dit roept de vraag op hoe data mee kan worden genomen door de tijd en toegankelijk kan blijven. En tot slot roept het gebruik van algoritmen door de Rijksoverheid nieuwe vragen op voor databeheer: welke data naast het algoritme, zoals als trainings- en testdata, moet op welk moment worden opgeslagen om verantwoording af te kunnen leggen in de toekomst?

Er zijn dus zeker de nodige uitdagingen en witte vlekken te benoemen vanuit technisch oogpunt, waarvoor (mogelijk) diverse oplossingen bestaan. Dat wil echter niet zeggen dat alle Rijksoverheidsorganisaties hiermee aan de slag moeten. Zij zullen eerst en vooral hun eigen situatie moeten analyseren om te begrijpen wat voor hen noodzakelijk is.

5.3 Normen en kaders

Hieronder beschrijven we kort de belangrijkste kaders en normen op meer technisch vlak. Het gaat daarbij niet om een uitputtende opsomming van ieder deelgebied, maar vooral om generieke kaders, die Rijksoverheidsorganisaties kunnen toepassen op hun eigen specifieke situatie⁶⁰. Het is bovendien goed daarbij te beseffen dat dit veld door de continue technische ontwikkelingen volop in beweging is, en dat het dus voor Rijksorganisaties zaak is de ontwikkelingen te blijven volgen.

5.3.1 DUTO-eisen en *archiving by design*

De eisen rondom duurzame toegankelijkheid (DUTO-eisen) van het Nationaal Archief zijn een *programma van eisen voor duurzame toegankelijkheid van de informatie in de informatiesystemen van overheidsorganisaties*⁶¹. Dit generieke programma van eisen kan door overheidsorganisaties – na een vertaling naar hun eigen specifieke context – worden gebruikt bij de aanschaf en inrichting van informatiesystemen. Daarmee kunnen zij voldoen aan de toetsbare eisen die de Archiefregeling (art. 16) stelt met betrekking tot te bewaren informatie, maar hebben ze ook meer grip op te vernietigen informatie die langere tijd bewaard moet worden. De DUTO-eisen hebben voorts nog de status van norm: ze zijn niet verplicht, maar wel sterk aan te bevelen. Overheidsorganisaties kunnen aan de slag gaan met deze eisen door middel van een zogenaamde DUTO-scan die wordt uitgevoerd door het Nationaal Archief.

Het principe ‘archiving by design’ sluit aan bij de DUTO-eisen. Volgens dit principe dient het beheer van informatie zoveel mogelijk ‘by design’ te worden georganiseerd. Dat wil zeggen dat informatiesystemen bij bouw, aanschaf of aanpassing zo worden ingericht dat de data duurzaam toegankelijk is, of de data uiteindelijk wordt overgebracht naar archiveringssysteem (als een e-depot of een DMS) of binnen het werksysteem zelf toegankelijk blijft⁶². Dit vereist ook samenwerking van archiveringsexperts met bijvoorbeeld architecten en functioneel ontwerpers, en biedt de mogelijkheid om samen op te trekken met experts op het gebied van privacy, beveiliging of bijvoorbeeld open data.

⁵⁹ <https://www.nationaalarchief.nl/archiveren/kennisbank/handreiking-digitaal-vernietigen>

⁶⁰ NORA biedt ook een goed overzicht van kader voor duurzame toegankelijkheid:

https://www.noraonline.nl/wiki/Wat_zijn_de_kaders_voor_Duurzame_Toegankelijkheid%3F

⁶¹ <https://www.nationaalarchief.nl/archiveren/duto>

⁶² <https://www.nationaalarchief.nl/archiveren/kennisbank/archiveren-by-design>

5.3.2 FAIR en NORA gegevensmanagement

FAIR is een afkorting voor *Findable* (vindbaar), *Accessible* (toegankelijk), *Interoperable* (uitwisselbaar) en *Reusable* (herbruikbaar). FAIR is een set principes afkomstig uit de wetenschap, die beogen om de herbruikbaarheid van wetenschappelijke data (en daarmee de reproduceerbaarheid van onderzoek) te borgen⁶³. Deze algemene principes zijn gericht op de metadata die nodig zijn om deze principes te realiseren. Daarmee fungeert FAIR als een set van prominente algemene, internationale principes voor het beheer van gegevens, die ook deels bij de Nederlandse overheid ingang heeft gevonden. Naast FAIR zijn er ook andere principes, zoals TADA⁶⁴ en diverse principes van de Digital Society⁶⁵ (VSNU). NORA Gegevensmanagement bevat, zoals hierboven al aangegeven, ook de nodige handvatten voor de inrichting van databeheer, en sluit aan bij de FAIR-principes⁶⁶.

5.3.3 NEN(-ISO) normen

De volgende NEN en NEN-ISO normen bieden Rijksorganisaties handvatten om niet alleen het beheer van bestanden, maar ook van data in te richten. Daarin worden ook technische aspecten benoemd:

- NEN 2082 (2008): norm voor record management met functionele eisen die moeten worden gesteld aan software om te kunnen voldoen aan wet- en regelgeving (waaronder de Archiefwet).
- NEN-ISO 15489-1(2016): norm voor informatie- en archiefmanagement waarop organisaties hun informatiehuishouding kunnen baseren om te voldoen aan wet- en regelgeving.
- NEN-ISO 23081-1 (2017): norm voor processen voor informatie en archiefbeheer, metagegevens voor archiefbestanden.

5.3.4 Metadatamodellen

Metadata zijn essentieel voor duurzaam databeheer⁶⁷. Het toepassingsprofiel metagegevens Rijksoverheid (TP Rijk) en het toepassingsprofiel metagegevens Lokale Overheden (TMLO) worden momenteel doorontwikkeld naar een nieuw model: 'Metagegevens Duurzaam Toegankelijke Overheidsinformatie' (MDTO)⁶⁸. Hiermee zal een overheidsbreed metadatamodel ontstaan dat kaders gaat geven voor de metadata die Rijksoverheidsorganisaties moeten meegeven aan hun informatie. Dit zal mogelijk niet één op één toepasbaar zijn op databases, maar bieden wel een startpunt voor de metadatering hiervan.

5.3.5 Instrumenten voor waardering en selectie

De eerste vraag die voor databeheer moet worden beantwoord is welke data gedurende welke periode moet worden beheerd. Ook hiervoor bestaan instrumenten. Het Generiek Waarderingsmodel Rijksoverheid (GWR)⁶⁹ is een model waarmee Rijksoverheidsorganisaties hun eigen selectielijsten

⁶³ <https://www.go-fair.org/> Zie ook Marnix van Berchum en Marjan Grootveld, 'Het beheren van onderzoeksdata', Handboek Informatiewetenschap [IV B 475].

⁶⁴ <https://tada.city/>

⁶⁵ <https://www.thedigitalsociety.info/nl/over/dataprincipes/>

⁶⁶ <https://www.noraonline.nl/wiki/Gegevensmanagement>

⁶⁷ <https://www.nationaalarchief.nl/archiveren/kennisbank/metadata>

⁶⁸ <https://www.nationaalarchief.nl/archiveren/nieuws/van-tmlo-en-tp-rijk-naar-mdto>

⁶⁹ <https://www.nationaalarchief.nl/archiveren/kennisbank/veelgestelde-vragen-over-waarderen-en-selecteren#collapse-446>

kunnen ontwerpen. Het bevat onder meer voorbeelden van werkprocessen en suggesties voor de waardering van dergelijke processen. Daarmee is het niet alleen een goed uitgangspunt voor de waardering van bestanden en processen, maar ook voor gestructureerde data. De Handreiking waardering en selectie van het Nationaal Archief biedt Rijksoverheidsorganisaties eveneens nuttige handvatten hiervoor⁷⁰.

5.3.6 Algoritmes

Algoritmes krijgen momenteel veel aandacht en er is recent het nodige over geschreven⁷¹. Er bestaat echter geen norm voor het beheer van algoritmes⁷², hoewel dit voor verantwoordingsdoeleinden wel noodzakelijk is. Dit geldt met name voor algoritmes die door bijvoorbeeld geautomatiseerde verwerking of besluitvorming een grote (potentiële) maatschappelijke impact hebben. Het ontbreken van een dergelijke norm is dan ook te beschouwen als een 'witte vlek' in databeheer. De componenten die gearchiveerd kunnen worden zijn de keuzes bij de ontwikkeling van het algoritme, het algoritme zelf, de broncode, training- en testdata en de software die de broncode uitvoert. Het is voor toegankelijkheid wenselijk hier een handleiding bij te voegen zodat ook niet-experts dit kunnen bekijken⁷³.

5.3.7 Overige kaders

Voor de publicatie van data wordt op diverse niveaus gewerkt aan centrale plaatsen (websites) waarop publicatie van data georganiseerd is. Dit moet zowel in het werkproces als in de technische functionaliteit geregeld worden. Aangezien de technologische ontwikkeling van API's steeds verder gevorderd is, kan een organisatie kiezen voor automatische of handmatige publicatie⁷⁴. Voor dergelijke uitdagingen is het Kennis en Exploitatiecentrum Officiële Overheidspublicaties (KOOP) ingericht⁷⁵. KOOP biedt ook, via data.overheid.nl, hulpmiddelen voor het inventariseren van data⁷⁶.

Voor anonimisering en pseudonimisering heeft o.a. het NEN gewerkt aan een norm (NEN 7524). Standaarden bestaan er weliswaar, en in overvloed, voor bestanden⁷⁷, maar eigenlijk niet voor databases. Een uitzondering is de SIARD-standaard voor relationele databases⁷⁸.

⁷⁰ <https://www.nationaalarchief.nl/archiveren/kennisbank/handreiking-waardering-en-selectie>

⁷¹ Zie bijvoorbeeld Marijn Janssen en George Kuk, 'The challenges and limits of big data algorithms in technocratic governance', *Government Information Quarterly*, 33 (3), pp. 371-377, de Kamerbrief Waarborgen tegen risico's van data-analyses door de overheid van 8 oktober 2019, Autoriteit Persoonsgegevens, *Toezicht op AI & Algoritmes*, 17 februari 2020 en het recente rapport van de Algemene Rekenkamer, *Aandacht voor Algoritmes*, uit januari 2021.

⁷² Petra Helwig, 'Rekenen en rekenschap. Algoritmes en Archiefwet', *Tijdschrift voor Toezicht*, 2020, 1, pp. 54-59; Inspectie Overheidsinformatie en Erfgoed, Een dementerende overheid 2.0? Een analyse over de informatiehuishouding bij het Rijk; 15 jaar "na een dementerende overheid?", februari 2021, p. 13.

⁷³ <https://algoritmeregister.amsterdam.nl/>

⁷⁴ <https://data.overheid.nl/ondersteuning/data-publiceren/datasets-publiceren> ; zie ook <https://data.overheid.nl/ondersteuning/data-publiceren/handreiking-open-data>

⁷⁵ <https://www.koopoverheid.nl/>

⁷⁶ <https://data.overheid.nl/ondersteuning/data-publiceren/inventarisaties>; zie ook <https://www.open-overheid.nl/open-overheid/inventariseren-kun-je-leren/>

⁷⁷ Zie de website van Forum Standaardisatie (<https://www.forumstandaardisatie.nl>) voor een overzicht van verplichte en aanbevolen standaarden.

⁷⁸ <https://dans.knaw.nl/nl/over/diensten/easy/toelichting-data-deponeren/voor-het-deponeren/bestandsformaten/siard>

5.4 Samenvatting

De technische kant van databeheer kent de nodige uitdagingen, zoals het beheer van relationele databases, logging, tijdreizen en vernietiging. Hiervoor bestaan er en der standaarden en normen. In de praktijk is de techniek volgend ten aanzien van de juridische en organisatorische eisen: pas als die goed zijn ingevuld is duidelijk wat er technisch nodig is om in een geconstateerde behoefte te voorzien.

Het gaat er bij techniek dus om een nadere invulling te geven aan datamanagement. Hiervoor bieden de DUTO-eisen, de FAIR-principes en NORA, net als de NEN-ISO normen en de bestaande metadatamodelen nuttige eerste handvatten. Het overkoepelende uitgangspunt hierbij is 'archiving by design', waarbij de gebruikte systemen aan de voorkant zo worden ingericht dat duurzaam databeheer ook daadwerkelijk mogelijk is. Belangrijk uitgangspunt bij dit alles is de selectielijst als onmisbaar basisinstrument voor waardering en selectie van data.

Niet alle technische randvoorwaarden voor databeheer zijn ingevuld. Het lijkt erop dat er momenteel niet goed is voorzien in conserverings- en vernietigingsfunctionaliteit voor data, mogelijk omdat deze aspecten buiten de standaard databeeprocessen vallen. Een opvallende technische uitdaging is het beheer van relationele databases. Ook de archivering van algoritmes is een 'witte vlek' waarvoor nog geen norm is ontwikkeld.

Technisch is er wel degelijk veel mogelijk, maar de wereld van duurzame toegankelijkheid heeft zich nog niet altijd doorvertaald in technische oplossingen voor duurzaam databeheer. Een reden hiervoor is waarschijnlijk dat de eerste stap nog niet overal is gezet, en dat is de bewustwording van het feit dat data ook onder de bestaande juridische kaders valt, dat daar dus organisatorisch het één en ander voor moet worden geregeld en dat dat zich uiteindelijk ook moet doorvertalen in technische oplossingen.

6. Casus: databeheer bij het CBS

6.1.1 Inleiding

Het Centraal Bureau voor de Statistiek (CBS) verwerkt grote hoeveelheden data over de meest uiteenlopende onderwerpen. Hierdoor is databeheer de *core business* van het CBS. De organisatie heeft dan ook veel kennis van en ervaring met dit onderwerp en is een interessante casus in het kader van deze verkenning. In dit afsluitende hoofdstuk beschrijven we daarom op hoofdlijnen hoe het CBS databeheer heeft georganiseerd en welke lessen hieruit zijn te trekken voor andere organisaties binnen de Rijksoverheid.

6.1.2 Organisatie van het databeheer

Op grond van de Wet op het Centraal bureau voor de statistiek mag het CBS voor statistische doeleinden veel en veelsoortige data van uiteenlopende overheidsorganisaties verwerken en beheren. Het CBS heeft als taak het maken van statistieken op basis van deze data. De data komt binnen op basis van bijv. (online) enquêtes en vragenlijsten of vanuit registers en administraties van organisaties als DUO en de Belastingdienst, vanuit basisregistraties als de BRP, vanuit de loonaangifteketen of bijvoorbeeld vanuit de sociale zekerheidsinformatie van gemeenten. In totaal gaat het om ongeveer tweehondervijftig registers.

Deze ruwe data wordt na binnenkomst verwerkt tot statistieken. Het CBS schrijft geen aanleverstandaard voor aan toeleveranciers van data, maar standaardiseert deze zelf. Het vervaardigen van de statistieken is verdeeld over drie verschillende directies, al naar gelang het type data, met daarnaast medewerkers voor ondersteuning en publicatie. Binnen de directies komt de data terecht bij verschillende afdelingen. De verschillende afdelingshoofden zijn proceseigenaren, en in die hoedanigheid verantwoordelijk voor 'hun' statistiek, maar ook voor de toegang tot de data.

De ruwe data wordt gepseudonimiseerd, gemetadateerd, geaggregeerd, geharmoniseerd en gekoppeld en er worden bewerkingen op uitgevoerd, zoals imputatie en het maken van afleidingen. Daarbij wordt het pseudonimiseren om veiligheidsredenen alleen door een select groepje gespecialiseerde medewerkers uitgevoerd. Voordat van gepseudonimiseerde data geanonimiseerde output wordt gemaakt in de vorm van statistieken en (computerleesbare) open data en deze output wordt gepubliceerd, wordt een controle uitgevoerd op de herleidbaarheid naar individuele personen. Het waarborgen van de privacy is zodoende een essentieel en integraal onderdeel van het beheer en de verwerking van de data door het CBS, van ontvangst tot en met publicatie. Het bewerken van sociaaleconomische en ruimtelijke data gebeurt binnen het Stelsel van Sociaal-Statistische Bestanden (SSB), dat een breed hergebruik van deze data mogelijk maakt⁷⁹. Bij al deze processen vinden controles plaats die zoveel mogelijk zijn geautomatiseerd, maar ook deels worden uitgevoerd door enkele statistisch analisten. Versiebeheer is essentieel voor het CBS, en bij iedere processtap vindt metadatering plaats. Deze metadatering is verplicht op het hoogste niveau, door de DG, en op de naleving van deze plicht wordt toegezien op basis van 'comply or explain'.

⁷⁹ <https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/korte-onderzoeksbeschrijvingen/stelsel-van-sociaal-statistische-bestanden--ssb-->

Het resultaat van de verschillende bewerkingen in de laatste fase vóór het anonimiseren is het echte 'goud' van het CBS: microdata. Dit is data van zeer hoge kwaliteit, die de basis vormt voor bijvoorbeeld onderzoek, maar ook voor de reguliere output in de vorm van open data, Statlinetabellen en andere publicatievormen. De schatkamer waarin dit goud wordt bewaard is het Data Service Centrum (DSC)⁸⁰, het centrale data-archief waarin de microdata wordt beheerd. Vroeger gebeurde dit lokaal, met alle beperkingen van dien voor de datakwaliteit en toegang; door de centralisering van het beheer worden deze ondervangen. Naast opslag en toekomstvaste borging biedt het DSC faciliteiten als toegangsbeheer, rechtenbeheer en een zoekfunctie op basis van de metadata, waardoor het fungeert als het centrale distributiepunt. De output (statistieken) wordt opgeslagen in een publiek toegankelijke website waar de statistieken en open data voor onbepaalde tijd voor iedereen kosteloos toegankelijk zijn. Ook deze informatie wordt (dagelijks) gearchiveerd, maar deze archivering vindt buiten het DSC plaats.

Technisch gezien heeft het CBS metadatering zoveel mogelijk geautomatiseerd. Voor het beheer van langlopende datareeksen wordt, al naar gelang de statistiek, regelmatig een complete 'afslag' opgeslagen in het DSC, waarbij het hele bestand opnieuw op wordt geslagen; tussentijds worden alleen de wijzigingen opgeslagen. Om data door de tijd heen mee te kunnen nemen, en dus te conserveren, wordt de data, samen met de metadata, in een eenvoudig formaat (CSV of ASCII) opgeslagen, dat ook na lange tijd nog goed uit te lezen is. Zo is weliswaar een enorm archief ontstaan in de loop van de tijd, maar is de totale omvang van de data-opslag, met zo'n 10 TB voor alleen het DSC, te overzien.

6.1.3 *Lessons learned*

Uit de manier waarop het CBS databeheer heeft georganiseerd, is een aantal interessante lessen te trekken voor andere organisaties:

1. De **selectielijst** als de basis: deze lijst is de basis voor het databeheer, omdat de organisatie hiermee de cruciale vraag beantwoordt welke data zij eigenlijk wil bewaren. Dat maakt het beantwoorden van de vraag hoe je komt tot een passend beheer eenvoudiger;
2. **Afspraken over eigenaarschap** van de data: deze afspraken creëren verantwoordelijkheid én duidelijkheid over de toegang en gebruiksrechten;
3. **Centrale opslag** (bij het CBS in het DSC): dit voorkomt dat data lokaal op verschillende manieren wordt beheerd door verschillende medewerkers en garandeert de duurzame toegankelijkheid ervan;
4. **Verplichte metadatering**: het werk van metadateren is niet het leukste, en degenen die er de 'lasten' van hebben zijn niet per se degenen die er ook de 'lusten' van hebben; de verplichtstelling van metadatering op het hoogste niveau binnen de organisatie (door de DG) en daarbij bijbehorende controle op de naleving (volgens het regime *comply or explain*) hebben een zeer grote positieve impact op de kwaliteit van de metadata. Adequate metadatering maakt het bovendien mogelijk in de grote hoeveelheid data heel precies te zoeken op thema, onderwerp of zelfs variabele;
5. **Cultuur**: veel werknemers van het CBS beschikken van nature over de precisie die vereist is voor adequaat databeheer en privacybescherming zit in de 'genen' van de organisatie.

⁸⁰ <https://www.cbs.nl/nl-nl/corporate/2016/20/data-service-centrum-cbs-maakt-hergebruik-van-data-eenvoudig>

Hierbij valt op dat deze verschillende lessen niet zozeer betrekking hebben op de technische aspecten van het beheer, maar vooral op de manier waarop dit beheer is ingebed in de organisatie en is gelinkt aan de werkprocessen en aan de medewerkers zelf.

Bijlage A Bestudeerde documentatie⁸¹

Juridische documenten

#	Naam
1	Grondwet (Gw)
2	Archiefwet (Aw)
3	Archiefbesluit
4	Archiefregeling
5	Algemene Verordening Gegevensbescherming (AVG)
6	Wet Openbaarheid van bestuur (Wob)
7	Richtlijn 2003/98/EG van het Europees parlement en de Raad van de Europese Unie van 17 november 2003 inzake hergebruik van overheidsinformatie
8	Richtlijn 2013/37/EU van het Europees Parlement en de Raad van 26 juni 2013 tot wijziging van Richtlijn 2003/98/EG inzake het hergebruik van overheidsinformatie
9	Wet hergebruik overheidsinformatie (Who)
10	Memorie van Toelichting Archiefwet
11	Memorie van Toelichting Wet openbaarheid van bestuur
12	Memorie van Toelichting Wet hergebruik overheidsinformatie
13	ECLI NL RvS 2019 899

Achtergronddocumentatie

#	Documentnaam	Datum
1	<i>Verslag parlementaire ondervragingscommissie Kinderopvangtoeslag. Ongekend onrecht</i>	17 december 2020
2	Kabinetsreactie rapport 'Ongekend onrecht' (Kamerbrief)	15 januari 2021
3	Waarborgen tegen risico's van data-analyses door de overheid (Kamerbrief)	8 oktober 2019
4	RDDI, <i>Samen op weg naar duurzaam toegankelijke overheidsinformatie. Meerjarenplan verbetering informatiehuishouding Rijksoverheid. Actualisering ten behoeve van 2020</i>	Januari 2020
5	Reframing studio en Future Lab, <i>De informatiehuishouding van het Rijk in 2030</i>	April 2020
6	Petra Helwig, 'Rekenen en rekenschap. Algoritmes en Archiefwet', <i>Tijdschrift voor Toezicht</i> , 2020, 1, pp. 54-59	2020

⁸¹ Het navolgende overzicht van de gebruikte achtergronddocumentatie is ingedeeld naar de wijze waarop deze in de tekst is aangehaald: hetzij als publicatie (waarbij wet- en regelgeving voor de duidelijkheid apart is genoemd), hetzij als website.

7	Inspectie Overheidsinformatie en Erfgoed, <i>Een dementerende overheid 2.0? Een analyse over de informatiehuishouding bij het Rijk; 15 jaar "na een dementerende overheid?"</i>	Februari 2021
8	Inspectie Overheidsinformatie en Erfgoed, <i>Landelijke voorzieningen en WOZ en Adressen en Gebouwen</i>	April 2020
9	Europese Commissie, <i>Ethische richtsnoeren voor betrouwbare KI</i>	2019
10	DAMA International, <i>Data Management Body Of Knowledge</i> , 2e ed.	2017
11	Datagovernancemodel Rijksoverheid (kenniswiel en selfassessment)	2018
12	Algemene Rekenkamer, <i>Aandacht voor Algoritmes</i>	januari 2021
13	Algemene Rekenkamer, <i>Machineleesbare gegevensbestanden; archivering en beheer bij het Rijk</i>	1991
14	Autoriteit Persoonsgegevens, <i>Toezicht op AI & Algoritmes</i>	17 februari 2020
15	Marijn Janssen en George Kuk, 'The challenges and limits of big data algorithms in technocratic governance', <i>Government Information Quarterly</i> , 33 (3), pp. 371-377.	2016
16	Marnix van Berchum en Marjan Grootveld, 'Het beheren van onderzoeksdata', <i>Handboek Informatiewetenschap [IV B 475]</i> , KNAW	December 2016
17	KVAN/BRAIN, <i>Weten of vergeten? handreiking voor het toepassen van de Algemene verordening gegevensbescherming in samenhang met de Archiefwet in de dagelijkse praktijk van het informatiebeheer bij de overheid.</i>	April 2020

Geraadpleegde websites

#	Organisatie	Link
1	KIA	https://kia.pleio.nl/blog/view/55812212/de-wettenkaart-online
2		https://kia.pleio.nl/files/view/55811587/overzicht-archieffregelingen-peter-brand.xlsx
3	LED	https://led.pleio.nl/groups/view/fbdb70e5-2c2d-4a62-b6b1-55fe53e61e5c/datagedreven-werken-community/files
4		https://led.pleio.nl/events/view/254fe0a0-bd1f-4ffe-8cce-92513598c4a4/chief-data-officer-een-waardevolle-positie-of-rol
5	NORA	https://www.noraonline.nl/wiki/Gegevensmanagement
6		https://www.noraonline.nl/wiki/Wettelijke_kaders_en_regelgeving_gegevensmanagement
7		https://www.noraonline.nl/wiki/Gegeven
8		https://www.noraonline.nl/wiki/Integraal_(begrip_gegevensmanagement)
9		https://www.noraonline.nl/wiki/Rollen_binnen_gegevensmanagement

10		https://www.noraonline.nl/wiki/Wat_zijn_de_kaders_voor_Duurzame_Toegankelijkheid%3F
11	GEMMA	https://www.gemmaonline.nl/index.php/GMT_Gegevensmanagement
12		https://www.gemmaonline.nl/index.php/GM_Gegevensfunctiehuis
13	Open overheid	https://open-overheid.nl/wp-content/uploads/2016/05/WEB_90943_BZK_Handleiding-Who-versie2.pdf
14		https://www.open-overheid.nl/open-overheid/inventariseren-kun-je-leren/
15	Rijksoverheid	https://www.rijksoverheid.nl/documenten/rapporten/2017/09/29/model-gegevensbeschermingseffectbeoordeling-rijksdienst-pia
16		https://www.rijksoverheid.nl/documenten/rapporten/2018/01/22/handleiding-algemene-verordening-gegevensbescherming
17	Utrecht Dataschool	https://dataschool.nl/deda/
18	Digitale overheid	https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/nieuwe-technologieen-data-en-ethiek/publieke-waarden/toolbox-voor-ethisch-verantwoorde-innovatie/
19	Forum Standaardisatie	https://www.forumstandaardisatie.nl
20	Nationaal Archief	https://www.nationaalarchief.nl/archiveren/kennisbank/wat-is-preservering
21		https://www.nationaalarchief.nl/archiveren/kennisbank/overzicht-van-begrippen
22		https://www.nationaalarchief.nl/archiveren/kennisbank/handreiking-digitaal-vernietigen
23		https://www.nationaalarchief.nl/archiveren/duto
24		https://www.nationaalarchief.nl/archiveren/kennisbank/archiveren-by-design
25		https://www.nationaalarchief.nl/archiveren/kennisbank/metadata
26		https://www.nationaalarchief.nl/archiveren/nieuws/van-tmlo-en-tp-rijk-naar-mdto
27		https://www.nationaalarchief.nl/archiveren/kennisbank/handreiking-waardering-en-selectie
28		https://www.nationaalarchief.nl/archiveren/kennisbank/veelgestelde-vragen-over-waarderen-en-selecteren#collapse-446
29	Digital Society (VSNU)	https://www.thedigitalsociety.info/nl/over/dataprincipes/
30	iBestuur	https://ibestuur.nl/podium/van-beleids-naar-datacyclus
31		https://ibestuur.nl/podium/geen-woorden-maar-data
32	Wikipedia	https://en.wikipedia.org/wiki/DIKW_pyramid
33	RDDI	https://www.informatiehuishouding.nl/Producten+%26+publicaties/richtlijnen/2018/05/01/baseline-informatiehuishouding-rijksoverheid

34		https://www.informatiehuishouding.nl/Producten+%26+publicaties/instrumenten/2019/09/17/kenniskaart-werken-met-overheidsinformatie
35		https://www.informatiehuishouding.nl/voormedewerkers
36	Gemeente Amsterdam	https://algoritmeregister.amsterdam.nl/
37	KOOP	https://www.koopoverheid.nl/
38		https://data.overheid.nl/ondersteuning/data-publiceren/datasets-publiceren
39		https://data.overheid.nl/ondersteuning/data-publiceren/handreiking-open-data
40		https://data.overheid.nl/ondersteuning/data-publiceren/inventarisaties
41	GO FAIR	https://www.go-fair.org/
42	TADA	https://tada.city/
43	DANS (KNAW/NWO)	https://dans.knaw.nl/nl/over/diensten/easy/toelichting-data-deponeren/voor-het-deponeren/bestandsformaten/siard
44	CBS	https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/korte-onderzoeksbeschrijvingen/stelsel-van-sociaal-statistische-bestanden-ssb--
45		https://www.cbs.nl/nl-nl/corporate/2016/20/data-service-centrum-cbs-maakt-hergebruik-van-data-eenvoudig

Bijlage B Geïnterviewde personen

Datum	Naam	Rol en organisatie
12-2-2021	Wim Stolk	Expertgroep gegevensmanagement NORA/GEMMA
18-2-2021	Ruben Dood, Leen Roosendaal, Rik van den Hurk	Centraal Bureau Statistiek
19-2-2021	Petra Helwig	Senior inspecteur Inspectie Overheidsinformatie en Erfgoed
19-2-2021	Erik Saaman	Strategisch adviseur Nationaal Archief