

Duurzaam digitaal databeheer bij de Rijksoverheid¹

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

¹ Deze twopager is een samenvatting van de bevindingen uit het rapport 'Duurzaam digitaal databeheer bij de Rijksoverheid: een verkenning'. Voor een vollediger beeld zij verwezen naar dit rapport.

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.