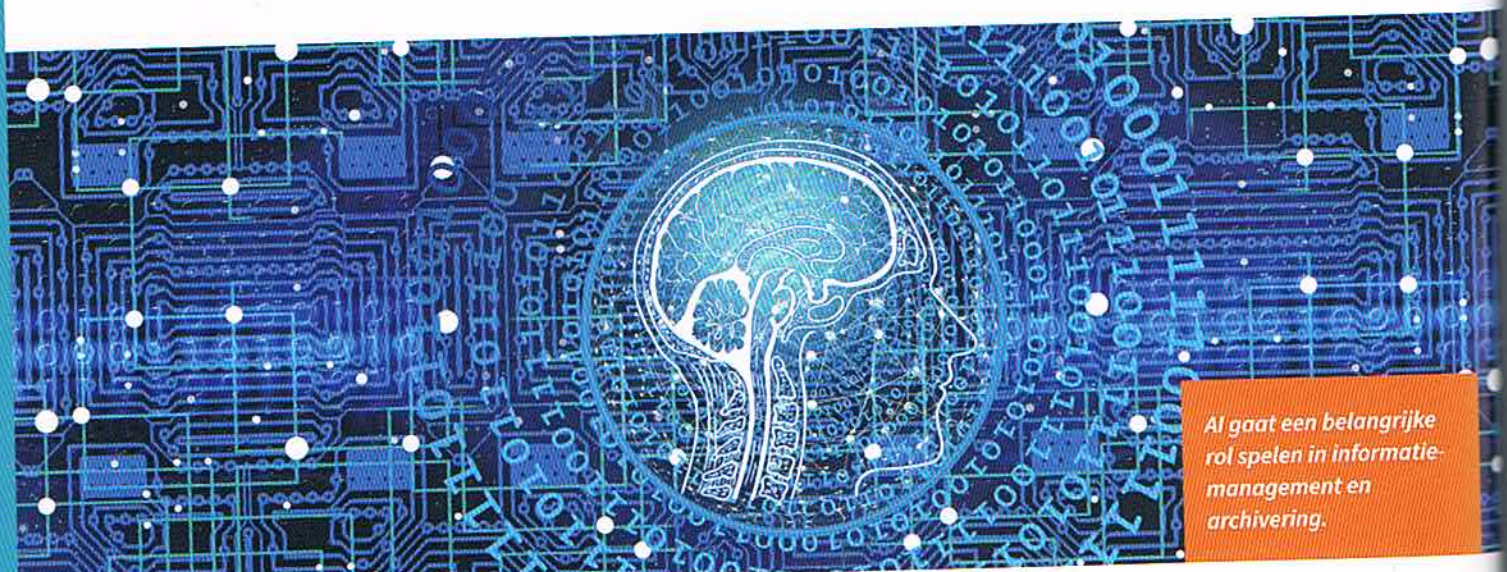


Een blik op de toekomst

Archivering in 2050



AI gaat een belangrijke rol spelen in informatie-management en archivering.

Hoe archiveren we in de toekomst? Hoe wordt informatie bewaard en ontsloten? Wat voor rol spelen automatisering en AI? Niemand weet wat de toekomst brengen zal, maar met de kennis van nu kunnen we wel een toekomstvisie schetsen over archivering in 2050.

In 2012 gaf ik in 'Archiving should be just like an Apple' een samenvatting van hoe archivering zich zou kunnen gaan ontwikkelen, namelijk het creëren van een omgeving waarin archivering volledig geautomatiseerd plaatsvindt.¹ Het idee was dat gebruikers geen of nauwelijks handelingen hoeven te verrichten om informatie contextueel te verwerken en op te slaan en dat archivering een achtergrondproces wordt. Medewerkers in organisaties doen hun werk, zonder zich af te vragen hoe informatie wordt opgeslagen en gearchiveerd. Ze kunnen erop vertrouwen dat informatie beschikbaar is op het benodigde moment (ongeacht wanneer dat is), voorzien van metadata om context, relevantie en kwaliteit ervan te kunnen beoordelen. Datzelfde geldt ook voor de informatie die als cultuurhistorisch erfgoed bewaard wordt in archiefbewaarplaatsen. Gebruikers van dat erfgoed moeten daarvan op dezelfde intuïtieve manier gebruik kunnen maken en die

archiefinformatie kunnen vinden en gebruiken, voorzien van alle context die nodig is om die informatie te kunnen interpreteren. Dat proces kan met allerlei tools en technieken worden ondersteund, zoals 'artificial intelligence' (AI) en 'machine learning' (ML). Informatie, ongeacht waar deze bewaard wordt, moet als het ware met een druk op de knop beschikbaar zijn, vergezeld van ontstaans-, gebruiks- en beheer(s)context.

In de 'Informatiemaatschappij van 2023' beschreven een aantal onderzoekers, ondernemers en beroepsbeoefenaars de in hun ogen onomkeerbare trends die de informatiemaatschappij van 2023 zouden kenmerken.² Veel trends werden gepresenteerd, waarvan sommige niet verwerkelijk zijn, andere wel. Er is (nog) geen 'universum van los zand', maar wel een 'stikken in gegevens'.³ De wereld draait (nog steeds) gewoon door, waarbij de digitalisering voortgaat en informatie zowel sociaal als politiek wordt.⁴ Ik onderscheidde zes trends, die normaal leken maar verre van gerealiseerd waren. Dat zijn ze ook nu niet. We zitten nog midden in de transitie van aanbodgestuurd naar vraaggestuurd. Informatiewijsheid staat er nog slechter voor dan tien jaar geleden. Informatie wordt nog steeds gezocht, al wordt dankzij zoektechnologische tools steeds meer informatie aangeboden die bij persoonlijke profielen en zoekopdrachten past. Of dat altijd de 'juiste'

Informatiewijsheid staat er nog slechter voor dan tien jaar geleden

informatie is, valt te betwijfelen. De informatiemaatschappij is contextueel, maar veel informatie wordt 'out-of-context' opgeslagen en bewaard. 'Linked data' zijn een feit en zijn via mobiele apparaten verbonden met mensen, vaak 'out-of-context' en onvolledig. En ten slotte: de 'open' informatiemaatschappij, traceerbaar en transparant, blijkt niet zo open en transparant, maar wel traceerbaar.⁵ Twee ontwikkelingen die ik zelf als bepalende trends zag werden nergens in de bijdragen genoemd: de milieuaspecten van informatietechnologie en de beperkende rol van wet- en regelgeving. Beide trends zijn de afgelopen jaren versterkt. De ontwikkelingen in het klimaat maken het noodzakelijk informatie-infrastructuren 'groen' te maken. De effecten van wet- en regelgeving (zoals AVG of Woo) beperken en versterken transparantie, toegankelijkheid en openbaarheid van informatie.⁶

De toekomst is altijd ongewis, maar op basis van de huidige stand van de informatietechnologie durf ik, voortbouwend op mijn schets van 2012, wel een blik te werpen op 2050. Dat is ver genoeg weg om zelf niet meer geconfronteerd te worden met het feit dat wat ik verwacht, niet blijkt te kloppen (en die kans is uiteraard groot).

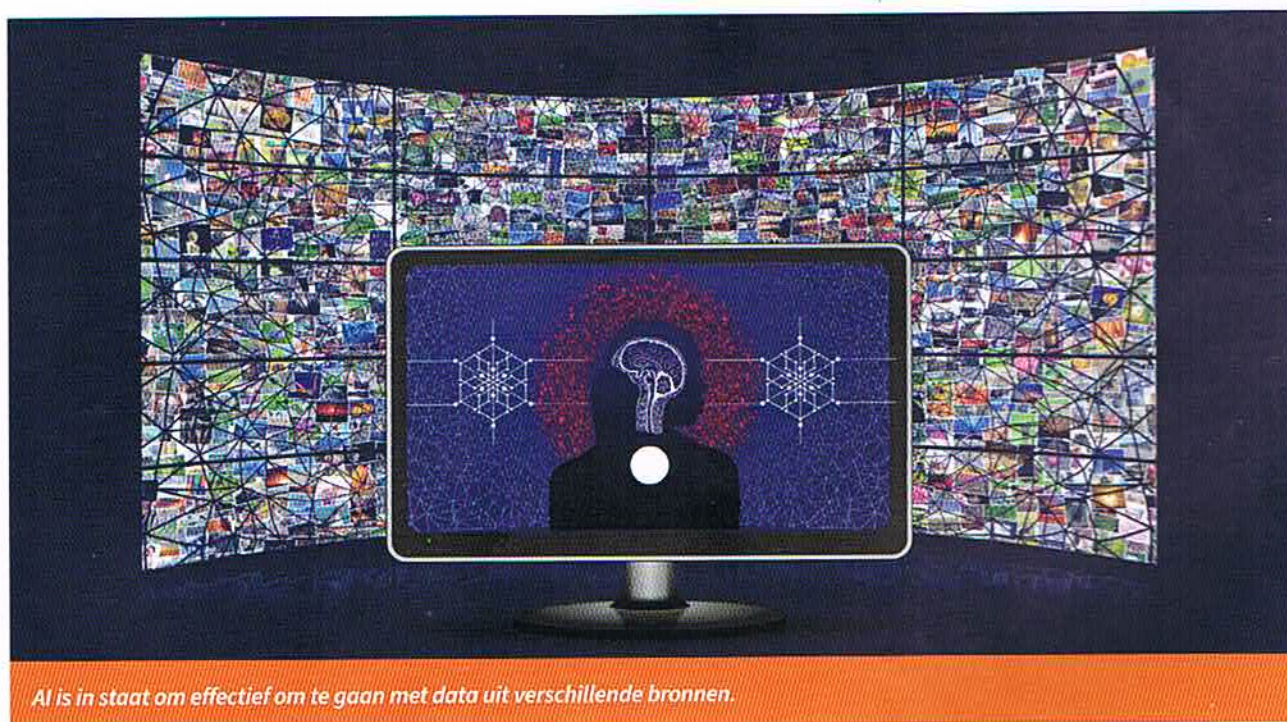
De volgende vooronderstellingen liggen ten grondslag aan mijn verwachtingen voor 2050:

1. Ik beschouw informatie als gegevens en gegevensobjecten. Wat voor archivering in aanmerking komt zijn de 'reproduceerbare verzamelingen samenhangende gegevens, die met gebruikmaking van een presentatievorm als eenheid worden gedragen, weergegeven of gecommuniceerd door middel van een medium, met de bedoeling om informatie over te dragen'.⁷ Gegevens kunnen op meerdere plekken zijn opgeslagen, mits de verzamelingen waar ze onderdeel van zijn (database records, documenten,

- GIS-datasets, CAD-tekeningen enzovoort) kunnen worden gereconstrueerd, gepresenteerd en ter beschikking gesteld zoals ze ooit waren op een eerder moment in de tijd.
2. Informatie wordt 'bewaard aan de bron', een al lang bestaand idee waarbij 'bron' op verschillende manieren is geïnterpreteerd (twintig jaar geleden bijvoorbeeld als de applicatie waarin de informatie werd gegenereerd). Ik definieer als 'bron' hier de archiefvormer, de organisatie die de informatie heeft gegenereerd in werkprocessen, en pas het toe op *alle* informatie die door een organisatie wordt gegenereerd. Niet enkel op die dynamische gegevenssets die onderdeel kunnen zijn van een 'onthefving tot overbrenging', zoals in de discussies over deze onthefving de insteek is.
 3. Mijn perspectief beslaat de gehele informatiewaardeketen, dus van het genereren tot de vernietiging en/of definitieve bewaring en publieke raadpleging van de informatie door onderzoekers (van genealogen tot academici), ongeacht de bewaarplaats van de informatie en ongeacht de plek van waaruit de informatie wordt benaderd.⁸
 4. Dit perspectief is gericht op de overheid. Hoewel het ook van toepassing is op bedrijfsorganisaties, is de specifieke invulling daarvan afwijkend, al is de rode draad identiek.

In 2050:

- A. Loggen medewerkers in, ongeacht de plaats waar ze verblijven, met hun mobiele werkunits, volledig versleuteld met post-kwantum cryptografie, bij het dashboardportaal van hun werkgever. Na een uitgevoerde authenticatie via zowel een gezichts- als een irisscan krijgen ze het op basis van hun autorisaties samengestelde persoonlijke dashboard gepresenteerd.
- B. Toont dat persoonlijke dashboard een overzicht van de openstaande werkzaamheden, veelal besluiten waarvan de organisatie in haar procesdefinities bepaald heeft dat



AI is in staat om effectief om te gaan met data uit verschillende bronnen.

- » menselijke beslissingen nodig zijn. Het werk van medewerkers bestaat uit het beoordelen van de dagelijkse rapportages van de AI over de afhandeling van zaken, het controleren van afwijkingen, de overleggen met betrokken burgers en bedrijven, het evalueren van gekozen strategieën en eventuele wijzigingen daarin communiceren met de AI-avatar voor verdere verwerking en het voorzien van bestuurders van informatie over politiek gevoelige zaken. De avatar van de AI is altijd ter beschikking voor ondersteuning en is voorzien van geavanceerde spraaktechnologie.
- C. Worden de meeste bedrijfsprocessen volledig automatisch afgehandeld op basis van door AI aangedreven 'business rules' in apps gespecialiseerd in de afhandeling van specifieke en complexe processen. Meerdere organisaties kunnen in deze processen worden betrokken en apps kunnen met behulp van beveiligde verbindingen met andere procesapps communiceren. De AI maakt gebruik van kennisbanken waarin per proces en processtap alle specificaties, vereisten en contextdata zijn vastgelegd en die tijdens de procesafhandeling automatisch worden toegewezen. Elke afwijking van de gedefinieerde werkstromen wordt aan drie menselijke beoordelaars toegewezen (eventueel in meerdere organisaties) die ieder afzonderlijk een beslissing nemen. De AI (eventueel samen met AI's van andere organisaties) bepaalt op basis van de antwoorden en de redenen die daarvoor zijn gegeven de mogelijke gevolgen, bepaalt welke beslissing genomen wordt, bespreekt die met de beoordelaars en verwerkt de afwijking, indien toegestaan, automatisch in de bestaande definities. ML zorgt ervoor dat de algoritmen van de lokale AI en de gebruikte apps voortdurend beter worden in het afhandelen van complexe processen. De AI biedt de aanpassing of aanvulling van de business rules en de redenen daarvoor aan de centrale kennisbank (van Rijk, VNG, Unie van Waterschappen) aan om de standaardmodellen (waar nodig) aan te passen. Beleids- en overlegprocessen krijgen automatische werkstromen, waarbij beleidsmedewerkers geassisteerd worden door de AI. De AI-avatar kan, in gesprek met de medewerkers, de te volgen procedure aanpassen en (interne en externe) deelnemers toevoegen en voorzien van de juiste autorisaties.
- D. Bevatten de centrale kennisbanken alle gestandaardiseerde datamodellen van procesdefinities, metadata, business rules, selectielijsten, archiefmodellen voor langdurige toegang en dergelijke, die gebruikt worden in de lokale

De consequentie van deze interpretatie van 'bewaren aan de bron' is dat het bestaande archiefbestel de mogelijkheden van de technologie niet overleeft



Steeds vaker zullen AI-avatars gerichte ondersteuning bieden bij informatievraagstukken.

- kennisbanken als richtlijn voor de specifieke datamodellen van een organisatie. De centrale kennisbanken worden door de lokale AI gebruikt bij het actueel houden van de lokale modellen, bij het beoordelen van de rechtmatigheid van lokale aanpassingen in procesdefinities en bij het bepalen of beslissingen op afwijkingen in hun gevolgen geoorloofd zijn.
- E. Heeft iedere organisatie een met quantumtechnologie beveiligde infrastructuur, gebaseerd op een Hybrid Storage Area Network (25e generatie, HSAN₂₅), waarin alle voorzieningen zijn getroffen voor beveiligde opslag en communicatie van gegevens en gegevensobjecten. HSAN₂₅ is zowel op premise als in de cloud, gefaciliteerd door de ICT-tak van de centrale organisaties (als Rijk, VNG en Unie van Waterschappen). De AI bepaalt op basis van centrale richtlijnen waar welke informatie wordt opgeslagen. Hierbij wordt gebruikgemaakt van meerdere commerciële partijen als leveranciers van het HSAN₂₅ en van datacenters waarin de hosting plaatsvindt. De datacenters functioneren op 'groene' energie, waarbij waterstof de voornaamste bron is, maar zon, wind en water ook bijdragen leveren. Alle opslag is redundant en voorzien van twee synchronisaties en drie mirrors, die in geval van calamiteiten doorwerken garanderen. In het HSAN₂₅ is ook de 'Digital Vault' gerealiseerd die dient voor die informatie die blijvend bewaard moet blijven. De vijf gebruikte datacenters zijn, gezien de klimaatproblematiek, gesitueerd in die gebieden in Europa waar de stijging van de zeespiegel geen rol van betekenis speelt.
- F. Is zoeken vinden geworden. Iedere medewerker vindt via het persoonlijke dashboard alle informatie die nodig is voor het uitvoeren van werk, inclusief alle gegevens over de context van de informatie en mogelijke relaties met andere (beschikbare) informatie. De AI-avatar kan, in overleg met de medewerker en rekening houdend met diens autorisaties, andere informatie aanleveren of suggereren. De avatar kan in voorkomende gevallen ook zelf antwoord geven.
- G. Vindt informatiemanagement bijna volledig geautomatiseerd plaats op basis van de processen van de informatiewaardeketen. Informatie wordt van intake tot vernietiging of definitieve bewaring voorzien van context. Alleen de ter zake doende persoonsgegevens worden vastgelegd (zichtbaar voor geautoriseerde medewerkers). De archiveringsprocessen

Informatiemanagement en archivering gaan naar volledige automatisering toe

worden uitgevoerd (de vernietigingstermijn (gebaseerd op procesresultaat) bepaalt het moment van vernietiging, en bestandsformaten worden, daar waar dat kan, omgezet naar duurzame formaten (gedefinieerd in formaat libraries als Pronom₂₀₅₀). Informatie wordt openbaar gemaakt op basis van de Wet open overheid (2041). Vernietigings- en overbrengingsprocessen worden geautomatiseerd uitgevoerd, waarbij enkel de archivaris een goedkeuring moet geven of uitzonderingen mag bepalen op basis van input van de AI-avatar over eventuele hotspots. De automatische 'overbrenging' naar de Digital Vault is voorzien van de voor publicatie noodzakelijke formaten en te publiceren context. De archivaris krijgt de beheerautorisaties toegewezen. Tweemaal per jaar wordt het informatiemanagement geïnspecteerd door een externe Audit AI, gebaseerd op een kennisbank van de Centrale Inspectie voor Informatiebeheer van het ministerie van Informatie en Informatie-infrastructuur, zoals voorgeschreven door de Informatiewet (2039). 'Oude' archieven zijn gedigitaliseerd. Enkel het erfgoedmateriaal dat dateert van voor 1850 wordt in origine bewaard, volgens alle criteria van goed beheer en in centrale archiefbewaarplaatsen, toegankelijk voor onderzoekers op afspraak. Al dit materiaal is digitaal beschikbaar via de Digital Vault.

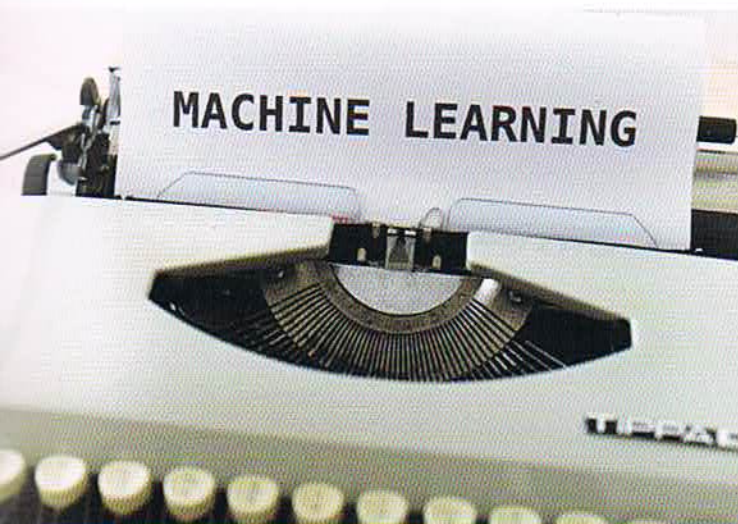
H. Bevat de Digital Vault alle te bewaren informatie, voorzien van alle benodigde context. Deze Vault is toegankelijk voor medewerkers binnen de organisatie. De Digital Vault wordt door het HSN₂₅ optimaal beveiligd. Een openbare

'mirror' van de Vault waarborgt de toegang voor het publiek. Deze mirror is doorzoekbaar en voorzien van een AI-avatar, die bezoekers ontvangt en helpt met het vinden van informatie. Deze mirror is vanuit de site van de organisatie, maar ook vanuit generieke en publieke erfgoed-sites benaderbaar. De informatie wordt gepresenteerd in oorspronkelijke vorm, geëmuleerd indien nodig, voorzien van transcripties en vertalingen in modern Nederlands door tools als Transkribus₆₈ of Translate₂₀₅₀. Zoekmachines beschikken over OCR-indexen van alle archieven (inclusief context) en maken het mogelijk informatie te raadplegen, eventueel uit meerdere Vaults tegelijkertijd.

Een deel van dit perspectief zou nu al gerealiseerd kunnen worden in basale vorm op basis van wat de huidige technologie biedt. Mijn perspectief betekent dat er niet langer afzonderlijke organen zijn voor het bewaren van archieven. De consequentie van deze interpretatie van 'bewaren aan de bron' is dat het bestaande archiefbestel de mogelijkheden van de technologie niet overleeft. Bovenstaand perspectief verandert echter nauwelijks als uitgegaan wordt van een perspectief vanuit het huidige bestel. Het impliceert een andere inrichting van het proces- en infrastructuur-landschap, maar blijft qua visie overeind. Informatiemanagement en archivering gaan naar volledige automatisering toe en AI zal daarin een belangrijke rol gaan spelen. AI kan ik het, zoals bij ieder perspectief, mis hebben. Maar ook dan draait de wereld gewoon door... |

NOTEN

- 1 G.J. van Bussel, 'Archiving should be just like an Apple™', en acht andere, nuttige (?) stellingen, HvA Publicaties (Amsterdam 2012), p. 41.
- 2 G.J. van Bussel (red.), *De informatiemaatschappij van 2023. Perspectieven op de nabije toekomst*, Hogeschool van Amsterdam/GEA (2013).
- 3 R. Knoop, 'Blade Runner. Het neerdalend stof van een catastrofe', in: G.J. van Bussel (red.), *De informatiemaatschappij van 2023. Perspectieven op de nabije toekomst*, Hogeschool van Amsterdam/GEA (2013), p. 12-23.
- 4 J. Mackenzie Owen, 'De wereld draait gewoon door', in: G.J. van Bussel (red.), *De informatiemaatschappij van 2023. Perspectieven op de nabije toekomst*, Hogeschool van Amsterdam/GEA (2013), p. 24-33.
- 5 G.J. van Bussel, 'Introductie', in: G.J. van Bussel (red.), *De informatiemaatschappij van 2023. Perspectieven op de nabije toekomst*, Hogeschool van Amsterdam/GEA (2013), p. 9-10.
- 6 Van Bussel, 'Introductie', p. 10. Ook: G.J. van Bussel, N. Smit, en J. van de Pas (2015), 'Digital Archiving, Green IT and Environment. Deleting data to manage critical effects of the Data Deluge', *Electronic Journal of Information Systems Evaluation* (2013), Vol. 18, no. 2, p. 187-198.
- 7 G.J. van Bussel, *Documenten onder controle. Optimaliseer uw informatievoorziening*, Kluwer, (Deventer 2008), p. 18.
- 8 G.J. van Bussel, 'The theoretical framework of the 'Archive-as-Is'. An organization-oriented view on archives. Part I. Setting the stage: enterprise information management and archival theories. Part II. An exploration of the 'Archive-as-Is' framework', in: F. Smit, A. Glaudemans, and R. Jonker (eds.), *Archives in Liquid Times*, SAP ('s-Gravenhage 2017), p. 16-41, p. 42-71.



Dankzij machine learning gaat software beter presteren door te leren van data.