

Data archiving in een digitale wereld

Routekaart voor de archivering van gestructureerde gegevens

Drs. Jeroen Tegelaar, drs. Peter Kuiters en drs. Jan Geurtsen

Data archiving wordt onterecht vaak alleen gezien als een lastige en tijdrovende bezigheid die na afhandeling van een proces of taak dient te worden uitgevoerd, een noodzakelijk kwaad waar men zich eens per jaar mee bezig moet houden. In dit artikel wordt data archiving juist beschouwd als iets wat strategische meerwaarde kan bieden en daarmee meer aandacht en waardering verdient. De misverstanden rondom dit thema worden weggenomen en na het vaststellen van een definitie volgt een overzicht van de voordelen. Vervolgens zal worden ingegaan op de verschillende oplossingsrichtingen die beschikbaar zijn voor de implementatie, waarbij ook aandacht zal zijn voor de specifieke criteria bij de keuze voor een applicatie of tool.



Drs. J.A.C. Tegelaar
is senior manager
bij KPMG IT Advisory.
tegelaar.jeroen@kpmg.nl



Drs. P. Kuiters
is manager bij
KPMG IT Advisory.
kuiters.peter@kpmg.nl



Drs. J.M.B. Geurtsen
is junior adviseur
bij KPMG IT Advisory.
geurtsen.jan@kpmg.nl

Inleiding

Binnen het vakgebied van datamanagement neemt archivering (in dit artikel data archiving genoemd) als thema een bijzondere positie in. Ten onrechte vaak bestempeld als de fase die het einde van een proces inluidt, is archivering bij uitstek het terrein waar gestructureerde data (lees: data opgeslagen en geordend in databases) en ongestructureerde data (lees: e-mail, documenten, social media content, etc.) elkaar in een ideale situatie zodanig aanvullen dat Enterprise Data Management kan worden bereikt: de situatie waarin alle informatie op het juiste moment voor de juiste persoon voor de juiste gelegenheid beschikbaar en toegankelijk is.

In de praktijk is dit echter meestal niet het geval, doordat achterstallig onderhoud op met name digitaal geboren data een dergelijke situatie onmogelijk maakt. Opmerkelijk genoeg zijn de meeste organisaties op de hoogte

1. In een artikel uit 2004 met de titel 'De tijd van aanmodderen is voorbij' (zie [Glaso4]) wordt betoogd dat organisaties zich onvoldoende realiseren dat de gestelde eisen aan papieren documenten en gerelateerde beheermaatregelen in een digitale wereld niet meer opgaan, omdat de kenmerken van digitale gegevens wezenlijk verschillen van de kenmerken van gegevens vastgelegd op papier. Ook het feit dat de distributie van data veel meer verloopt via informele kanalen als e-mail, web-based applicaties en internet, in plaats van via de vroegere formele (en nog steeds bestaande) kanalen als postkamers en secretariaten, vraagt om een radicaal andere aanpak van de archivering van data.

van dit probleem, maar ondernemen ze geen actie, ook al weten ze dat dit op den duur tot ontoelaatbare risico's en hoge reconstructiekosten kan leiden. De oorzaak hiervan is eenvoudig: verantwoordelijke managers weten niet hoe men archivering moet aanpakken omdat de methodieken uit het verleden voor het beheer van het papieren archief niet meer van toepassing zijn¹.

Zo zien IT-beheerorganisaties data archiving vooral als een interessant selectie-instrument om een onderscheid te maken tussen data die direct nut heeft voor de dagelijkse bedrijfsvoering en overige data. De eerste groep dient altijd direct beschikbaar en toegankelijk te zijn. De rest kan ergens anders offline of in de cloud een stuk goedkoper worden opgeslagen, tot de afgesproken bewaartermijn is verstreken.

Dat archivering van digitale data een andere insteek behoeft, die niet alleen de problemen van deze tijd oplost, maar ook nieuwe kansen biedt wordt helaas nog onvoldoende beseft. Om duidelijk te krijgen wat data archiving in een digitale wereld kan opleveren, zal eerst duidelijk moeten worden wat we onder dit begrip verstaan en welke misverstanden over dit onderwerp de ronde doen.

Enkele misverstanden over data archiving

Over data archiving bestaan nogal wat misvattingen. Het meest voorkomende misverstand is dat bewaren gelijk staat aan archiveren. Dit is niet correct. Simpelweg bewaren is het onbeheerd opslaan, waarbij vaak verschillende methoden en opslagmedia worden gebruikt. Archiveren gaat veel verder. De waarde van data wordt bepaald, relevante gegevens worden geselecteerd, bewaartermijnen worden geïdentificeerd en maatregelen ter bescherming van gegevens opgesteld op basis van locatie van de data en verwachte toegang. Archiveren betekent namelijk dat gegevens nog steeds actief worden beheerd, met als voorwaarde dat de gegevens zelf niet meer veranderen.

Het grote verschil met 'bewaren' zit in de gestructureerde aanpak en toepassing van beleidsmaatregelen. Data archiving is de voorlaatste stap in de levenscyclus van data, de volgende en laatste stap is vernietiging.

Het tweede grote misverstand is dat activiteiten rond data-archivering pas dienen te worden gestart aan het einde van het proces. Vooral voor ongestructureerde data geldt dat hoe eerder bekend is voor welke doeleinden en voor welk proces data wordt gecreëerd en geraadpleegd, hoe eerder de waarde van de data bekend is. Het gevolg is dat deze beter kan worden beschermd en tegelijkertijd kan worden gedistribueerd aan de juiste personen gedurende zijn levenscyclus. Ook is eerder duidelijk welke bewaartermijnen dienen te worden gehanteerd, waarmee ook de IT-beheerorganisatie is geholpen in haar streven de kosten voor databeheer zo laag mogelijk te houden.

Een derde misverstand is de gelijkstelling van data archivering met het proces van het maken van back-ups. Het hebben van een back-up is een noodzakelijke continuïteitsmaatregel en is gericht op herstel na een systeemfalen. De aandacht gaat in feite uit naar het beschikbaar zijn van de applicatie waarmee een bedrijfsproces wordt ondersteund na een eventuele calamiteit. Bij archivering wordt gekeken naar de gegevens en komt de applicatie op de tweede plaats. Gegevens overleven vaak de applicatie waar ze in worden beheerd of worden zoals eerder al gesteld, omwille van het maken van ruimte gemigreerd naar een alternatieve omgeving. Dit kan zijn naar een andere goedkopere locatie of een nieuwe applicatie. Back-ups van data kunnen een maatregel van archivering zijn, maar dat is in de praktijk meestal niet de reden voor het maken van een back-up.

Data archiving is in onze visie dan ook niet alleen het archiveren van digitale data, maar het geheel aan processen, procedures, activiteiten, maatregelen en middelen die als doelstelling kennen data voor een vastgestelde periode zodanig onveranderbaar vast te leggen, dat toegankelijkheid en beschikbaarheid voor de juiste persoon op het gewenste moment kunnen worden gerealiseerd. In dit artikel richten wij ons op de archivering van reeds digitaal geboren data in het IT-domein, alhoewel wij ons wel degelijk realiseren dat de archivering van fysieke documenten als onderdeel van data archiving kan worden gezien. De focus ligt verder op de problematiek van archivering van gestructureerde data, in een later artikel zullen wij ingaan op de archivering van ongestructureerde data, omdat deze qua aanpak en aanleiding wezenlijk verschilt en andere ondersteuning behoeft.

Archiveren staat niet gelijk aan bewaren

Nut en noodzaak van data archiving

Nu duidelijk is geworden wat data archiving omvat, kijken we nu naar de toegevoegde waarde die data archiving kan opleveren.

Data archiving als een middel om kosten te besparen op serverruimte is zojuist al genoemd als aanleiding, maar er zijn nog veel meer redenen die een aanleiding kunnen zijn om data archiving hoog op de agenda te zetten. De invoering van het Nieuwe Werken is een eerste belangrijke aanleiding, waarbij de wensen van plaats- en tijdonafhankelijk werken alleen kunnen worden gerealiseerd als informatie op elk gewenst moment beschikbaar en toegankelijk is in de juiste vorm. Hiervoor dient het digitaal archief uiteraard ook op orde te zijn, anders vinden medewerkers niet wat ze zoeken. Een andere vaak genoemde aanleiding is de uitfasering van applicaties, waarbij de organisatie de data in de applicaties graag nog wil behouden of moet behouden, bijvoorbeeld als gevolg van bewaarplicht. Het moeten voldoen aan de wettelijke verplichtingen rondom het bewaren van informatie zoals opgelegd door de belastingdienst, maar ook het op tijd vernietigen van persoonsgegevens op basis van de Wet bescherming persoonsgegevens (Wbp) is vaak reden om te starten met data archiving. Data archiving komt soms hoog op de agenda te staan, omdat een organisatie is geconfronteerd met de hoge kosten van een verloren rechtszaak omdat bepaalde bewijsvoering niet meer terug te vinden was. Tot slot het noemen waard: data archiving is beter voor

het milieu. Dit is simpelweg doordat er minder hardware nodig is voor beheer als er minder data wordt bewaard buiten de verstreken bewaartermijnen. Dit vertaalt zich in minder stroomverbruik en daarmee een lagere belasting voor het milieu.

Organisaties en de waarde van data

Bovenstaande aanleidingen kunnen worden geclusterd in vier verschillende groepen van redenen, zoals grafisch is weergegeven en toegelicht in figuur 1².

Hierbij dient direct te worden opgemerkt, dat de meeste organisaties niet beseffen dat de genoemde redenen vaak verschillende belangen dienen. Belangen die nog wel eens tegenstrijdig zijn ten opzichte van elkaar.

Zo zal een organisatie vanuit het principe van verantwoording in meerdere of mindere mate willen voldoen aan verschillende eisen voortvloeiend uit wet- en regelgeving, maar gaat men in dit streven al te vaak voorbij aan de impact die bepaalde maatregelen kunnen hebben op de dagelijkse bedrijfsvoering en vice versa. Het gebruik en beheer van persoonsgegevens is hiervan een mooi voorbeeld. Vanuit het perspectief van bedrijfsvoering kunnen persoonsgegevens van klanten waardevolle inzichten leveren voor marketing- en salesdoeleinden. Ook persoonsgegevens van de eigen medewerkers, bijvoorbeeld verkregen uit meerdere bronnen als medewerkerstevredenheids-

2. Het artikel van Eric Ketelaar, emeritus hoogleraar archiefwetenschap aan de Universiteit van Amsterdam, getiteld 'De waarde(n) van archieven', noemt deze bovengenoemde belangen van archivering. Verder stelt hij dat velen onterecht het bedrijfsleven als 'wezens zonder geheugen' zien, naar een analogie van Marc Bloch. Het archief zou worden gezien als een bewaarplaats van oude informatie die niet relevant is voor de realiteit van de *competitive marketplace*. Ketelaar merkt echter fijntjes op dat grote organisaties met een lange staat van dienst als Coca-Cola en Shell wel degelijk de waarde van historische archieven inzien, in de vorm van inkomsten uit hergebruik van oud advertentiemateriaal en als bron voor de verdediging van patinten en merken. Shell beheert zelfs zorgvuldig de rapporten over alle olieboringen sinds 1890: er moet op worden teruggegrepen wanneer een reeds lang verlaten put een nieuwe kans krijgt dankzij nieuwe exploratietechnologie. Zie ook: [Keteo4].



Figuur 1. Waarde van archieven.

Archiveren is goed voor het milieu

onderzoeken of functioneringsgesprekken, kunnen iets vertellen over de bestaande bedrijfscultuur en de kwaliteit van de bedrijfsvoering. Vanuit het perspectief van verantwoording dienen deze gegevens echter strikt te worden gebruikt voor de doeleinden waarvoor ze zijn opgevraagd en na gebruik dienen ze feitelijk per omgaande onherstelbaar te worden vernietigd. Als de verantwoordelijke afdelingen voor deze gegevens zich niet realiseren welk effect het kan hebben deze gegevens ongeoorloofd te bewaren zonder hiervoor maatregelen te hebben getroffen, dan loopt de organisatie plotseling het risico op boetes vanuit de verantwoordelijke toezichthouder (in casu: het CBP). Belangrijker nog dan de boete (die absoluut gesproken voor veel organisaties van bescheiden omvang is) is de potentiële imagoschade. Deze is veelal een veelvoud van de kosten van de initiële boete. Het opzetten en inbedden van een goede archiveringsstrategie en deze toetsen aan de verschillende eisen die de organisatie stelt aan informatie, is daarom zeker geen overbodige luxe.

Scenario's voor data archiving

Wanneer gegevens niet beschikbaar zijn heeft dit een negatief effect op de dagelijkse bedrijfsvoering en maakt een organisatie kosten die niet worden gedekt door omzet. Hoe hoog die kosten zijn hangt af van de impact die het ontbreken van deze specifieke gegevens heeft, ofwel de waarde van de gegevens. De eerste stap bij data archiving is daarom altijd een analyse van de waarde van de gegevens. Hier komt de strategische waarde van data archiving naar voren. Het tijdelijk en soms zelf permanent wegvallen van deze gegevens vormt namelijk altijd een onacceptabel risico dat moet worden gemitigeerd door een combinatie van oplossingen, maatregelen en procedures. Om die reden is de bovengenoemde analyse in de regel vaak een strategische risicoanalyse op de gegevens. Een binaire blik op gegevens, waarbij alleen wordt bepaald of gegevens wel of niet moeten worden bewaard, is te beperkt en leidt tot problemen in de afweging van belangen. Zonder de kosten van het bewaren en beheren hierbij mee te nemen, kan dit leiden tot willekeur en hoge kosten voor beheer omdat men voor de zekerheid alles wil bewaren. Door risico's als een factor mee te wegen (op een schaal van 1 tot 10 bijvoorbeeld) kan een afweging worden gemaakt tussen risico's en kosten voor de mitigatie. Dit geeft een meer genuanceerde en gebalanceerde uitwerking die de basis vormt van een archiveringsstrategie, na de risicoanalyse de tweede noodzakelijke stap binnen het domein van data archiving. Uiteraard kan het nog steeds absoluut noodzakelijk zijn

om gegevens te bewaren, bijvoorbeeld in het geval van wettelijke voorschriften, maar dit zal altijd meer uitzondering dan regel zijn en dergelijke gegevens dienen dan ook een uitzonderingspositie te krijgen.

Het helder hebben van de gegevens, de waarde daarvan, het gebruik en de benodigde bescherming is een belangrijke stap in de voorbereiding van de daadwerkelijke archivering van data. Met dit inzicht kunnen we nu de stap zetten naar de implementatie. Voor we ingaan op de aspecten die relevant zijn voor de keuze van de specifieke oplossing, beschrijven we in deze paragraaf een aantal verschillende scenario's.

Voor data archiving is een grote hoeveelheid methodieken en oplossingen beschikbaar in de markt. Om in deze veelvoud aan keuzes de weg te vinden, hebben wij een indeling gemaakt over twee assen. Ten eerste het object van archivering, applicatie of data. Ten tweede het vraagstuk van centralisatie versus decentralisatie. Hieronder worden deze twee indelingen kort toegelicht.

Bij een applicatiegerichte strategie worden de bestaande applicaties in feite intact gelaten. Dit kan in de meest eenvoudige vorm de continuatie van de bestaande server betekenen. Wanneer het object van archivering de data betreft worden de relevante gegevens geselecteerd en in een andere vorm buiten de applicatie opgeslagen. Decentrale oplossingen hebben een tactisch karakter en gaan uit van meerdere losstaande datasets, bijvoorbeeld door per applicatie een aparte archiveringsoplossing te kiezen. Bij centralisatie worden de gegevens juist centraal in één oplossing bewaard, beter bekend onder de term 'datawarehousing'.

Wanneer beide kenmerken tegen elkaar worden afgezet ontstaat een matrix van vier basis-archiveringsstrategieën (zie figuur 2)³. Deze zijn:

1. continuatie; waarbij de bestaande applicaties onafhankelijk van elkaar in grotendeels ongewijzigde vorm worden gecontinueerd;
2. virtualisatie; waarbij de bestaande applicaties worden omgezet naar virtuele servers, en ze een pool van beschikbare hardware delen;
3. tooling; waarbij de relevante gegevens in een applicatie-specifieke tool worden opgeslagen;
4. managed; waarbij de gegevens in een centrale tool worden opgeslagen en beheerd.

3. Gebaseerd op een combinatie van de bestaande methodologieën van BiSL, ITIL en ISO-15489 en ervaringen in de praktijk.

Het is van belang hierbij op te merken dat deze matrix geen vaste set van vier varianten vormt, maar een continuüm is voor de verschillende oplossingen. In het vervolg van deze paragraaf worden de vier scenario's beschreven en de voor- en nadelen aangegeven, waarbij we ons tot de hoofdlijnen zullen beperken.

Scenario 1. Continuatie

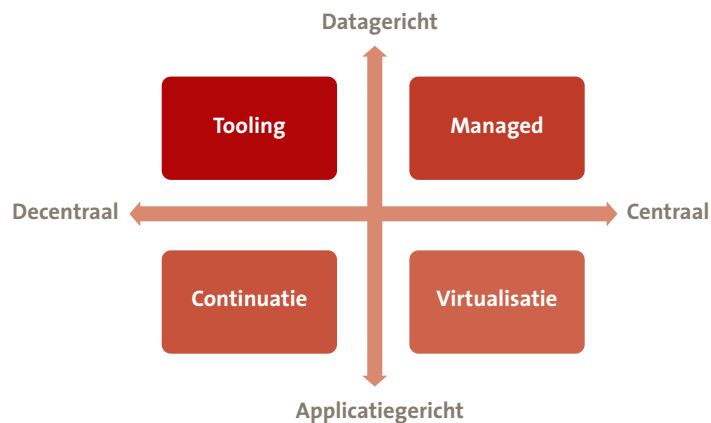
Het continuatiescenario is het meest eenvoudig en wordt ook veelvuldig toegepast. De implementatie is eenvoudig; in feite wordt er niets gedaan. De bestaande applicaties met onderliggende server blijven gewoon bestaan. In onze ervaring is de voornaamste reden dat deze strategie veel wordt toegepast het ontbreken van een aanpak voor archivering. Zodra na de implementatie van een nieuwe applicatie de oude moet worden uitgezet, ontstaat twijfel of alle relevante gegevens, bijvoorbeeld historische, wel zijn overgezet of anders opgevangen. Om geen enkel risico te lopen, laat men de oude applicaties naast de nieuwe applicaties intact.

Het laten functioneren van de bestaande servers heeft eigenlijk alleen nadelen. Ten eerste de technische beheerkosten; de bestaande omgeving is veelal groot opgezet en voor een archiefsysteem is dit veelal niet nodig. Bij een kostenstructuur op basis van capaciteit kan al snel te veel betaald worden. Tevens kunnen deze slapende systemen een nadeel vormen bij een datacenterconsolidatie, waarbij een grote hoeveelheid legacy servers de kosten flink kan opdrijven.

Een ander nadeel is dat hardware aan veroudering onderhevig is of dat door certificerings- en onderhoudseisen van de leverancier de configuratie van het systeem moet worden aangepast. Een jaar of zes is wel het maximum dat hardware meegaat, voor data is dat eerder het minimum. Het wordt daardoor noodzakelijk systeemverbeteringen door te voeren, zowel op het gebied van hardware, operating system als applicatiesoftware. Voor kleinere applicaties, vooral wanneer die op een gevirtualiseerde server staan, is deze impact kleiner dan bij grote applicaties, zoals ERP-, CRM-, HRM- en SCM-systemen. Bij deze laatste groep zal ook de applicatiekennis moeten worden bijgehouden door de eigen beheerorganisatie.

Scenario 2. Virtualisatie

Virtualisatie is een goede manier om de kosten voor beheer van weinig gebruikte applicaties te reduceren. De kern van de oplossing is dat de archive-



Figuur 2. Vier basis-archiveringsstrategieën.

ringsapplicaties voor het grootste deel van de tijd kunnen worden gedeactiveerd. Hierdoor is er alleen opslag nodig van de zogenaamde image en door het uitzetten van de omgevingen worden ook de meeste beheerkosten weggenomen. Zowel de opslag als de pool van virtuele hardware wordt centraal geregeld, daarmee is deze strategie fundamenteel anders dan de continuatiestrategie.

In dit scenario wordt de benodigde hardware betrokken uit de beschikbare overcapaciteit in het eigen datacenter of uit de cloud. Vooral voor grotere applicaties kan cloud een goede optie zijn, er is slechts tijdelijk een flink stuk extra capaciteit nodig. Uiteraard zijn er bij cloud aanvullende randvoorwaarden die dienen te worden ingevuld, zoals met betrekking tot de locatie van de opslag van gegevens. Dat kan echter per geval worden besloten en als alternatief kunnen lokale servers worden ingezet.

Door het up-to-date houden van de virtualisatieomgeving wordt de afhankelijkheid van oude hardware weggenomen en neemt de flexibiliteit in het datacenter toe. Gegeven de opslagduur van data kan het voorkomen dat de gekozen configuratie buiten het onderhoudscontract van de leverancier valt en een upgrade toch noodzakelijk is. Het moge duidelijk zijn dat dit scenario alleen geschikt is wanneer de toegang tot de gearchiveerde applicaties slechts zeer incidenteel is. Net als in het vorige scenario zal de kennis over de gearchiveerde applicaties moeten worden bijgehouden; met bewaartermijnen die kunnen oplopen tot tien jaar, kan dit een lastige opgave zijn en haaks staan op rationalisatie-initiatieven voor de rest van het applicatielandschap.

Continuatie van de bestaande situatie is de meest kostbare en risicovolle optie

Bij Managed is het leren van één archiveringsoplossing voldoende om toegang te krijgen tot alle relevante gegevens

Scenario 3. Tooling

Bij het inzetten van tooling ligt de nadruk op de relevante gegevens en worden deze overgezet naar applicatiespecifieke tools. Het is te verwachten dat deze tool de applicatiecontext behoudt (look-and-feel en gegevensstructuur), zodat de gebruikersacceptatie wordt verhoogd. Voor een eenmalige of kleinschalige oplossing is dit een goede mogelijkheid en bij het ontbreken van een brede archiveringsstrategie ook vaak het resultaat van een selectie. Een belangrijk aspect bij dit scenario is dat hier een gegevensselectie gemaakt wordt. Immers, alleen de data in de tools wordt bewaard, de rest wordt verwijderd. Deze gegevensselectie is een nuttige, maar geen triviale aangelegenheid.

Een simpele vorm van tooling is de opslag van archiefdata in de vorm van rapportages. Uitgangspunt zijn wettelijke rapportageverplichtingen, bijvoorbeeld voor de belastingdienst. In deze aanpak worden de verplichte rapportages aangemaakt en in een houdbaar formaat opgeslagen, bijvoorbeeld PDF-a. Voor de bewaarplicht zoals opgelegd door de belastingdienst is dit voldoende (zie ook [Belaro]). Het is evident dat bij deze strategie er geen ad-hoc toegang tot de gegevens mogelijk is. De verdere bruikbaarheid van de op deze wijze opgeslagen data is dan ook zeer laag.

Ten opzichte van de applicatiegerichte oplossingen zijn de beheerkosten vaak significant lager. Dit komt doordat een flinke hoeveelheid van de gegevens wordt verwijderd, waardoor de storagekosten dalen. Een ander voordeel ligt in de ondersteunde hardware en software. Door te kiezen voor een separate oplossing en deze ook up-to-date te houden, blijven de gegevens in een gecertificeerde en door de leverancier ondersteunde configuratie. Het kunnen inzetten van moderne (virtuele) hardware is een bijkomend voordeel dat positief bijdraagt aan de continuïteit en flexibiliteit van het landschap.

Voor een bredere aanpak is tooling niet ideaal omdat de gegevens over meerdere archieftools zijn verspreid waardoor relaties tussen data moeilijk te achterhalen zijn. Ook moet kennis van verschillende archieftools up-to-date gehouden worden.

Scenario 4. Managed

Wanneer de gegevens worden gecentraliseerd in één oplossing komen we in een managed-strategie. In dit scenario is veelal sprake van een datamanagementstrategie waarbin-

nen de gehele data-lifecycle is beschreven en wordt nageleefd met behulp van de gekozen archiveringsoplossing. Het implementeren van een centrale oplossing is vaak het meest kostbaar. Dit komt doordat de oplossing zelden voor alle te archiveren applicaties hulpmiddelen biedt voor de archivering. De inspanning voor het gecontroleerd overzetten van de gegevens wordt daardoor vaak hoger.

Uiteraard moet in dit scenario een gegevensselectie worden uitgevoerd om te komen tot de te archiveren set van data. Het verlies van applicatiecontext kan nadelig werken voor de gebruikersacceptatie, maar daar staan twee dingen tegenover:

1. Alle data komt in hetzelfde systeem terecht en er hoeft maar één archiveringsoplossing te worden geleerd om toegang te krijgen tot alle relevante gegevens.
2. Het wordt mogelijk om relaties tussen data uit de verschillende systemen te leggen zonder gebruik te maken van systeemkoppelingen.

Vooraf het tweede punt kan voor grote voordelen zorgen bij de ontsluiting van de waarde van de gegevens, bijvoorbeeld door aansluiting op een datawarehouse. Op deze wijze kan data beter worden benut in het kader van hergebruik, of om trends te ontdekken in de data (data-mining). Doordat alle gegevens centraal worden beheerd, kan maximaal worden geprofiteerd van geavanceerde oplossingen waarbij de datapolicies automatisch worden afgedwongen. Te denken valt aan het automatisch verwijderen van gevoelige gegevens na het verstrijken van de bewaartermijn. Hiermee worden de arbeidskosten van het beheer teruggedrongen.

Relevante criteria bij de keuze van een (software)oplossing

Na het bepalen van de waarde van data en het verwachte gebruik dient vervolgens in de formulering van een archiveringsstrategie en de keuze voor een oplossing rekening te worden gehouden met de volgende aspecten:

- gebruikersacceptatie;
- veiligheid en continuïteit;
- functionaliteit;
- kosten.

Gebruikersacceptatie

Het niet accepteren van een archiveringsoplossing door de gebruikers is funest voor een succesvolle realisatie van een archiveringstraject. Oude applicaties worden niet opgeruimd of de historische data wordt niet verwijderd en/of gemigreerd uit de bestaande applicatie. In beide gevallen zullen de beoogde voordelen niet worden gehaald. Zonder hier uitputtend te willen zijn, noemen we de volgende aspecten die leiden tot een snellere acceptatie door gebruikers van strategie en oplossing:

- Behoud van applicatiecontext. De look-and-feel en gegevensstructuur van de gearchiveerde data is vergelijkbaar met de bronapplicatie. Dit aspect is vooral relevant voor een tactische archiveringsoplossing (tooling).
- Gebruikersvriendelijkheid. Een prettig te gebruiken oplossing is essentieel voor de gebruikersacceptatie, vooral wanneer de applicatiecontext wordt weggelaten. Het niet nodig zijn van zware opleidingstrajecten van medewerkers maakt de oplossing goedkoper.

Veiligheid en continuïteit

Zoals eerder aangegeven vertegenwoordigen de opgeslagen gegevens een waarde voor de organisatie. Een betrouwbare oplossing, waarbij de toegang tot de gegevens aansluit bij de rechten binnen de organisatie, is noodzakelijk om de waarde van de gegevens recht te doen; dit speelt vooral wanneer gegevens van meerdere systemen worden samengevoegd. Dit betreft niet alleen beveiliging tegen externe toegang, zoals persoonsgegevens, maar ook interne toegang waarbij medewerkers van verschillende juridische entiteiten geen toegang mogen hebben tot elkaars financiële gegevens. De continuïteit kan gewaarborgd worden door middel van gangbare back-upprocedures.

Functionaliteit

Elke applicatie biedt verschillende functionaliteiten en zo ook archiveringsoplossingen. Te denken valt aan de ad-hoc-uitvragingsmogelijkheden. Deze moeten aansluiten bij de dagelijkse praktijk van de organisatie en er dient een balans te bestaan tussen de inspanning voor en het resultaat van een uitvraging. Ook zaken als ondersteuning van gerelateerde ongestructureerde gegevens, zoals gescande leveranciersfacturen, hebben duidelijk een meerwaarde.

Kosten

Vrijwel elke denkbare combinatie van gebruikersacceptatie, veiligheid en functionaliteit is te realiseren. Een hogere score op de verschillende onderdelen drijft echter de investeringskosten op. Voor de afweging om te komen tot een passende oplossing kan hier weer gebruik worden gemaakt van de eerdergenoemde risicoanalyse.

Praktijkvoorbeeld: archivering ERP-systeem

Een organisatie verandert van leverancier van haar ERP-systeem. Bij de implementatie wordt gekozen voor het overzetten van de courante gegevens, de historische gegevens worden niet meegenomen. Deze gegevens dienen echter wel te worden bewaard in verband met wettelijke verplichtingen. Een archiveringsstrategie ontbreekt, maar het onderwerp staat wel hoog op de agenda.

Het behouden van de applicatieomgeving wordt door licentie, hardware en beheerkosten als onwenselijk beschouwd. Virtualisatie is niet mogelijk door de regelmatige benadering van de gegevens. Er is geen keuze voor een centrale tool, dus valt de keuze op een tactische oplossing voor deze ERP-omgeving (tooling-scenario). De kosten van de implementatie van deze tool liggen significant lager dan het behouden van het bestaande systeem.

Bij de selectie van de tool zijn de meest zwaarwegende argumenten:

- de implementatiekosten, waarbij meegeleverde data-migratietools zwaar meetellen;
- de gebruiksvriendelijkheid, waarbij behoud van applicatiecontext van groot belang is.

Op de punten veiligheid en functionaliteit bleek weinig onderscheidend vermogen tussen de bekeken oplossingen. De organisatie in kwestie kiest uiteindelijk voor het scenario van tooling, omdat dit voor haar de beste balans lijkt te zijn voor een goede inbedding bij de bestaande beheerorganisatie in relatie tot de relatief lage kosten voor migratie.

Data archiving dient als een strategisch voordeel te worden benaderd, wil de werkelijke toegevoegde waarde zichtbaar worden

Conclusies

Bovenstaande uiteenzetting leidt tot de volgende conclusies:

- Data archiving dient niet te worden gezien als een probleem of als een technische truc van het IT-domein, maar dient strategisch te worden benaderd om de werkelijke toegevoegde waarde zichtbaar te maken.
- De aanleiding voor data archiving mag dan vaak verschillend zijn, de principes achter data archiving zijn dat niet. Deze vallen altijd terug te leiden tot de noodzaak van een efficiënte bedrijfsvoering, verantwoording over de uitgevoerde werkzaamheden, hergebruik van informatie en historische doeleinden.
- Data archiving kan vanuit de applicatie of vanuit de data worden bekeken. Een analyse vooraf op de waarde van gegevens is echter wel raadzaam om de kosten van het project beheersbaar te houden.
- Vier strategische scenario's kunnen worden onderscheiden ten aanzien van data archiving. Het scenario van *continuatie* kan daarbij als het meest kostbare en risicovolle worden beschouwd. Het scenario van *virtualisatie* als het meest interessante vanuit een beheertechnisch oogpunt. Het scenario van *tooling* is de meest praktische oplossing. Tot slot het scenario van *managed*. Dit lijkt het meest waardevol op de lange termijn, maar brengt ook de grootste initiële investering met zich mee.
- Tijdens een keuze voor een oplossing dient altijd rekening te worden gehouden met gebruikersacceptatie, functionaliteit, veiligheid en businesscontinuïteit en uiteraard de kosten.

Literatuur

- [Bel10] Belastingdienst 2010, *Uw geautomatiseerde administratie en de fiscale bewaarplicht*.
- [Glas04] Boudien J. Glashouwer RE RI CISA en Jan Pasmooij RA RE RO, *De tijd van aanmodderen is voorbij*, Controllers Magazine 2004/3.
- [Keteo4] Prof. dr. F.C.J. Ketelaar, *De waarde(n) van archieven*, Archievenblad 108/2, maart 2004, p. 16-19.

Over de auteurs

- Drs. J.A.C. Tegelaar** is als senior manager werkzaam bij KPMG IT Advisory. Hij heeft een zeer ruime ervaring met archiveringstrajecten op strategisch, tactisch en operationeel vlak en is binnen de Service Line Enterprise Data Management het aanspreekpunt voor vragen op het terrein van Document Management en Records Management.
- Drs. P. Kuiters** werkt als manager bij KPMG IT Advisory. Hij heeft een ruime ervaring met het opstellen van diverse typen architecturen (van Enterprise tot IT) en het begeleiden van trajecten waarin de architectuur wordt onderzocht. Zijn klanten kenmerken zich veelal door uitdagingen bij de beheersing van complexe heterogene applicatielandschappen.
- Drs. J.M.B. Geurtsen** werkt als junior adviseur bij KPMG IT Advisory. Hij is de afgelopen jaren betrokken geweest bij trajecten van informatiebeveiliging en data archiving, met name in de financiële sector.