Archiveren in de cloud?

Inleiding

Enige tijd geleden stond er op de site van de Dutch cowboys een artikel over hoe je om moet gaan met informatiebeheer in de "Cloud"[1]. De auteur van het artikel was in principe niet negatief over de mogelijkheden van de Cloud, maar waarschuwde wel voor mogelijke problemen. Ook van de KPN is er een brochure beschikbaar waarin de mogelijkheden van werken in de Cloud worden aangeprezen. Een van de aanvankelijke deelnemers aan de pilot heeft zich teruggetrokken. Men wilde zich gaan bezinnen op de gratis mogelijkheden van archiveren via bijvoorbeeld Dropbox. Het Landelijk Dienstencentrum, verschillende bedrijven en het ADC getroosten zich veel moeite en inspanning om een betaalbare oplossing te vinden voor de grote uitdagingen die het digitaal archiveren met zich meebrengt. Daarom is de vraag natuurlijk terecht: is dat het wel waard? Immers wanneer je gratis gebruik kunt maken van diensten als Dropbox, Google Drive, Skydrive en iCloud, waarom zou je dan een behoorlijk bedrag per jaar neertellen voor een systeem dat ogenschijnlijk weinig meerwaarde lijkt te bieden? Daarom wil ik hier ingaan op de vraag: Is archiveren met behulp van gratis clouddiensten mogelijk?

Wat is de Cloud?

Dit begrip wordt veel gebruikt en kent verschillende associaties. In principe betekent het niet anders dan dat de verwerking en opslag van de gegevens via het internet gebeurt. Dat betekent dat je data niet op je eigen computer staat, maar ergens op servers in een datacentre. Wanneer we het hier over de “Cloud” hebben, bedoelen we in de eerste plaats de gratis clouddiensten zoals Dropbox, Google Drive, iCloud enz. Als je gebruik maakt van gespecialiseerde diensten en van een soort e-depot, werk je eigenlijk ook in de Cloud, want ook die gaan via een internetverbinding en werken met opslag elders op een serversysteem. Nu is werken in de Cloud niet per definitie fout. De Cloud biedt zeker voordelen boven lokale opslag van informatie. Waar je ook bent, je kunt met een computer en internet altijd bij je gegevens. Zo heb je een enorme uitbreiding van je mobiliteit en flexibiliteit gekregen. Daarnaast zijn de gegevens ook relatief veilig. Lokale rampen en computerproblemen kunnen je informatie niet meer deren. Ook zijn de diensten breed gedragen door de maatschappij, waardoor er veel mogelijkheden zijn voor het uitwisselen van gegevens en voor synchronisatie van verschillende onderdelen. Een laatste, niet te geringschatten, voordeel vormen de lage kosten die eraan verbonden zijn. Veel opslag op kleinere schaal is volkomen gratis. Daarnaast wordt programmatuur voor beheer van de informatie door deze diensten ook gratis aangeboden via internet. Het is niet verwonderlijk dat voor veel mensen de keuze dan snel gemaakt is. Er zijn echter ook enkele grote "maar-s".

Wat is archiveren?

Die maar-s hebben niet in de laatste plaats te maken met de vraag wat je onder archiveren verstaat. Dit is niet slechts het voor de langere termijn opslaan van documenten. Aan een archief, fysiek of digitaal, worden enkele basale eisen gesteld. De opgeslagen informatie moet zo worden bewaard dat het na een lange tijd nog steeds toegankelijk is. Dit betekent dat de informatie niet alleen zolang intact bewaard moet blijven, maar dat die ook toegankelijk of te wel binnen afzienbare tijd vindbaar en leesbaar moet zijn. Behalve deze hele basale eis gelden er ook nog andere belangrijke eisen. De informatie moet veilig zijn.

[1] Het artikel: Wat als het fout gaat met je clouddienst?: http://www.dutchcowboys.nl/online/30771

Dat wil zeggen veilig voor calamiteiten zoals natuurrampen en diefstal, maar ook veilig in de zin van privacybescherming. Niet iedereen heeft zomaar toegang tot alle documenten. Privé-gegevens van mensen moeten niet in onbevoegde handen vallen. Een andere eis is dat de authenticiteit en integriteit van de informatie moet kunnen worden gewaarborgd. Is het document wel dat wat het pretendeert te zijn? Zeker bij digitale informatie is het vrij gemakkelijk de integriteit van documenten te schaden. Om al deze belangrijke eisen in een digitale omgeving te waarborgen nemen de specialisten hun toevlucht tot het woord "uitdaging". Met andere woorden het is niet bepaald makkelijk om dit voor elkaar te krijgen. Hoe moeten die basale eisen: *duurzaamheid, toegankelijkheid en veiligheid* worden gerealiseerd in het digitaal archiefbeheer en wat betekent dit voor het gebruik van gratis clouddiensten?

Bewaartermijn

Een van de belangrijkste punten van goed archiefbeheer is dat de informatie duurzaam bewaard wordt. Het moet dus lang houdbaar zijn[2]. De opgeslagen documenten hebben een lange bewaartermijn. Dit is iets dat gratis clouddiensten niet kunnen garanderen. Want waar blijft je informatie wanneer de dienst wegvalt door faillissement of om een andere reden? Het is niet waarschijnlijk dat de gebruiker zijn data terug krijgt. Dit is geen ondenkbare situatie gezien de beoogde bewaartermijnen van onze documenten (100 jaar en meer). Ook al ogen die bedrijven nu gezond, het is niet te verwachten dat ze over meer dan honderd jaar nog steeds zullen bestaan. Het is daarentegen wel de bedoeling dat onze digitale archieven er dan nog zijn.

Behalve het wegvallen van de clouddienst is er ook nog het probleem van onzorgvuldig handelen van de gebruiker. Wat gebeurt er wanneer je per ongeluk bestanden wist. Zijn ze dan voorgoed weg of zijn er toch nog mogelijkheden om die terug te halen? De eis van duurzaamheid vraagt om een gezond bedrijf dat garanties kan bieden over de beschikbaarheid van de documenten.

Toegankelijkheid

Het onderdeel toegankelijkheid splits ik hier op in twee componenten: formaat en metadata.

Formaat

Behalve duurzaamheid van de documenten is ook de toegankelijkheid een belangrijk aspect van goed archiefbeheer. Immers moet een digitaal document niet alleen lang kunnen worden bewaard, maar moet het na bijvoorbeeld vijftig jaar ook nog goed te lezen zijn. Dit is vooral een softwarematig aspect en heeft voor een belangrijk deel te maken met het opslagformaat. Dus datgene wat er na de punt van een bestandsnaam komt te staan (.doc, .pdf, .jpg enz.). Binnen een periode van tien tot vijftien jaar kan de software al zover zijn doorontwikkeld dat oudere formaten niet meer te lezen zijn. De oplossing voor dit probleem is het gebruik van officieel goedgekeurde standaardformaten. Een veel gebruikt voorbeeld hiervan is PDF-A/1. Een andere goede mogelijkheid is het gebruik van open source standaarden. Dit betekent dat de broncode van gebruikte software op één of andere manier is vrijgegeven voor algemeen gebruik. De gratis clouddiensten kunnen ook hier geen garanties bieden. Zij hebben eigen formaten (zoals in Google Docs) of maken gebruik van de formaten van de gebruiker. Een goed digitaal archiefsysteem houdt rekening met deze ontwikkeling en zorgt actief voor het toegankelijk houden van de informatie.

Metadata

[2] In de archiefwereld spreekt men wel over “bewaren voor de eeuwigheid”.

Een ander aspect van toegankelijkheid betreft de metadata die de digitale informatie dienen te begeleiden. Metadata is informatie over de informatie en is bedoeld om de informatie te kunnen plaatsen in de context van het geheel aan documenten van het archief. Want ook al kun je een document prima lezen, wanneer je er helemaal geen informatie over hebt behalve de inhoud van het document kan het vaak nog erg lastig zijn om het document goed te kunnen plaatsen. Er blijven vragen over als: uit welk archief komt dit document? Wie heeft het geschreven? Welke versie is dit eigenlijk? Met welk programma is dit document gemaakt? Wanneer is het gemaakt? Is het door een of andere instantie goedgekeurd ? Enzovoort. Deze metadata dienen met de informatie waar ze bij horen te worden gearchiveerd. Nu is het zo dat bijna elk programma wel metadata opslaat bij de documenten die er mee worden gemaakt. Het is echter wel de bedoeling dat juist die metadata kunnen worden gemaakt die we er ook echt in willen hebben. Daarnaast moet het mogelijk zijn die metadata te archiveren en ze moeten ook met andere programma's dan het programma waarmee ze zijn gemaakt gelezen kunnen worden. De gratis clouddiensten bieden deze mogelijkheden niet.

Veiligheid

Hieronder vallen verschillende aspecten van het digitaal archiveren. Het gaat hier zowel om het beveiligen van de informatie als om het beveiligen tegen die informatie. Met andere woorden het gaat om het veilig stellen van de informatie tegen virussen, corruptie, diefstal en calamiteiten, maar ook om het beschermen van de privacy van personen waar het in de informatie om gaat.

Bescherming tegen calamiteiten

De bescherming van de informatie tegen calamiteiten zoals brand, overstroming en natuurrampen of oorlog is een zaak van het datacentre waar de documenten zijn opgeslagen. Bij gratis clouddiensten is het niet altijd even duidelijk waar deze centra zijn gevestigd. Is dit in een gebied met relatief weinig risico's op natuurrampen of politieke onlusten of staan die in landen die goedkoop zijn, maar op dit punt minder hoog scoren? Daarnaast is het belangrijk dat de centra een vestiging op een andere geografische locatie hebben staan die een soort schaduw-archief heeft van de eerste locatie voor het geval er met een van de locaties iets gebeurt. Men mag verwachten dat grote organisaties zoals Microsoft, Google, Apple en Dropbox op dit punt hun zaken wel geregeld hebben. Toch blijft het een nadeel dat er geen controle of inspraak van de gebruiker mogelijk is.

Bescherming tegen corruptie, diefstal en misbruik

Behalve voor natuurlijke calamiteiten moet men ook beducht zijn voor bedreiging van mensen. Te denken valt hierbij aan het aanbrengen van virussen, aan vervalsing van de informatie of het (opzettelijk) wijzigen ervan zonder dat men daar toe gerechtigd is. Ook kan je hierbij denken aan diefstal en misbruik van de informatie door derden. Datacentra zullen hun servers goed moeten beveiligen tegen hackers. Bij bovengenoemde grote organisaties zal dit meestal wel op orde zijn. Toch zijn er ook voorbeelden van grote bedrijven die in dit opzicht niet altijd even goed scoren: Microsoft bijvoorbeeld.

Je moet niet alleen letten op bedreigingen van buitenaf, maar ook op bedreigingen van binnenuit. Dus van de organisaties zelf. Immers van wie is de informatie die door de gebruikers op de servers van deze instellingen wordt gezet? Mr. V.A. de Pous heeft dit aan de hand van de gebruiksvoorwaarden bekeken voor de dienst Drive van Google. Hij komt tot een aantal opmerkelijke conclusies. In de eerste plaats erkent Google de intellectuele eigendomsrechten van de gebruiker. Dus van de inhoud die iemand op Drive plaatst blijft die

persoon intellectueel eigenaar. Dit lijkt duidelijk, maar hier zit al een addertje onder het gras. Jij bent weliswaar eigenaar van de intellectuele inhoud, maar niet van de fysieke kopieën en dergelijk van de klantbestanden van Google. Dit blijft met alles wat erop staat eigendom van Google.
Behalve de eigendomsrechten verdient ook het gebruiksrecht onze aandacht. Volgens de voorwaarden geeft de gebruiker Google en alle bedrijven waarmee Google werkt een riant gebruiksrecht van de content. Google en een onbekend aantal zakelijke partners mogen de informatie gebruiken voor het draaien, promoten en verbeteren van de cloudopslagdienst. De beperking dat het alleen ten dienste van het draaien, promoten en verbeteren van de clouddienst is biedt in feite geen enkele zekerheid. Wat Google verstaat onder het promoten en verbeteren van de clouddienst kan wel eens iets heel anders zijn dan wat de gebruiker hierbij voor ogen heeft. Of een kerk haar vertrouwelijke documenten aangaande pastorale en financiële zaken van haar gemeenteleden ten dienste wil stellen van Googles (ongetwijfeld goudkleurige) idee van het verbeteren van de clouddienst lijkt me zeer de vraag.
De digitale informatie die een gebruiker laat archiveren bij een e-depot moet zowel intellectueel als feitelijk het eigendom van de gebruiker blijven. Dit is juridisch te regelen via bijvoorbeeld een escrow-overeenkomst. De beheerder van een e-depot heeft zeker geen gebruiksrecht en kan in principe niet bij de informatie zelf.

Bescherming van de authenticiteit en integriteit van de informatie
Bij de veiligheid van documenten hoort ook het aspect van authenticiteit en integriteit. Bij authenticiteit gaat het om de vraag of je wel het echte document te pakken hebt en niet een vervalsing. Bij integriteit gaat het om de vraag of er met het document niet geknoeid is. Oftewel: is een document wat het zegt dat het is? Papieren documenten zijn natuurlijk ook te vervalsen, maar dat vraagt een zekere vaardigheid en moeite. Kerkelijke documenten zijn doorgaans niet zo waardevol dat ze hiervoor in aanmerking komen. Bij digitale documenten is er echter veel meer mogelijkheid om documenten ongezien te veranderen. De drempel om even snel iets te wijzigen in een document is hierdoor veel lager geworden. Zelfs het voorheen onveranderlijke formaat PDF is tegenwoordig vaak nog wel aan te passen. Een goed e-depot houdt hier rekening mee. Het systeem kan controleren of het document is gewijzigd ten opzichte van het moment dat werd ingevoerd. Wijzigingen kunnen worden getraceerd en indien nodig kan men vorige versies terughalen. Een gratis clouddienst zoals Dropbox kent dergelijke uitgebreide controles niet. Wijzigingen zijn makkelijk gemaakt. Vorige versies zijn niet meer terug te halen. Hier hoeft overigens niet altijd sprake te zijn van kwade opzet. Het is heel makkelijk om iets verkeerd op te slaan, want je kunt via Dropbox rechtstreeks in je document werken. Ook dingen per ongeluk wissen is mogelijk. Als je deze documenten al terug kunt halen is dit vaak lastig en je moet er niet te lang mee wachten. Bescherming van de authenticiteit en integriteit wordt niet alleen gewaarborgd door de controlemechanismen van het systeem, maar ook door zaken als autorisatie (zie hieronder).

Bescherming door goede autorisatie
Een aspect dat veel met het bovenstaande te maken heeft is de bescherming van de privacy via een goede autorisatie van degenen die werken met het systeem. De autorisatie bepaalt wie er toegang heeft tot welke stukken. Een goede autorisatie zorgt zowel voor bescherming van de informatie zelf als voor bescherming tegen onterecht gebruik van de informatie. Vertrouwelijkheid is voor kerken een belangrijk punt. In de kerkelijke praktijk gaat veel vertrouwelijke informatie om. Denk bijvoorbeeld aan pastorale en financiële gegevens van gemeenteleden. Dit is allemaal informatie waar zorgvuldig mee omgegaan moet worden. In kerken zijn er echter nogal wat verschillende groepen die op een of andere manier recht hebben op inzage en gebruik van een gedeelte van de informatiestroom. Hoe zorg je ervoor

dat alle personen toegang krijgen tot die stukken waar ze recht op hebben zonder dat ze daarbij en passant de verkeerde stukken tegenkomen? De oplossing zit hem in een goede autorisatiemodule op het systeem. Dit regelt welke mensen welke stukken mogen inzien. Het regelt ook wie er iets mag wijzigen. Tenslotte moet zo een systeem goed kunnen omgaan met een veelvuldig verloop van geautoriseerde personen. Het mag niet zo zijn dat iemand die al twee jaar geen ouderling is nog steeds bij de stukken van de huidige kerkenraad kan komen. De gratis clouddiensten hebben wel bepaalde mogelijkheden voor het delen van documenten en samenwerken, maar schieten uiteindelijk tekort.

Conclusie

Zijn de gratis clouddiensten een goed middel voor duurzaam archiefbeheer? Ik denk van niet. Deze diensten zijn erg handig voor persoonlijk gebruik en voor het uitwisselen van documenten over verschillende platforms, maar voor een duurzaam en verantwoord digitaal archiefbeheer zijn ze ongeschikt. Een goed systeem is duurder dan de meestal gratis[3] diensten die Dropbox, Google Drive c.s. aanbieden. Hierover wil ik nog kort iets zeggen. Deze kosten zouden niet moeten worden gezien als een lastige extra post op de begroting, maar moeten worden meegenomen in de begroting als standaard onderdeel van het “runnen van een kerk”. Dit gebeurt bijvoorbeeld ook al met de kosten van de software voor het ledenadministratieprogramma. Daarnaast is het zo dat de kosten aanzienlijk lager zouden kunnen uitkomen wanneer er meer kerken hetzelfde systeem zouden gebruiken. Tenslotte gaat het hier om systemen met een hoog professioneel en gespecialiseerd karakter. De prijs-kwaliteit verhouding is zo bezien vaak erg gunstig.

Goed digitaal archiefbeheer is niet alleen een kwestie van ICT. Een goede archivaris kijkt verder dan de huidige trend op digitaal gebied. Hij/zij is geïnteresseerd in de periode waarin de huidige techniek inmiddels weer verouderd is. Is de informatie dan nog steeds beschikbaar? Kunnen we het dan nog steeds lezen, zien, horen? Hij/zij probeert nu al - voor zover dat mogelijk is - in te spelen op die latere situatie. Hadden we vijftien jaar geleden al onze tekstbestanden omgezet naar PDF, dan hadden we het nu een stuk makkelijker. Hadden we destijds bedacht dat de floppy disc wellicht vervangen zou worden door andere opslagmedia dan hadden we nu misschien nog de stukken van de periode 1995-2002 tot onze beschikking. Laten we over twintig jaar niet hoeven zeggen: hadden we destijds ons hele archief maar niet aan Dropbox toevertrouwd. En: bij goed digitaal archiefbeheer gaat het uiteindelijk niet om techniek, maar om beleid. De techniek heeft hierin slechts een dienende taak. Wanneer het beleid niet goed is, kan zelfs de beste techniek geen uitkomst bieden.

[3] Meestal gratis betekent hier: het is gebaseerd op freemium modellen. Het begint gratis, maar na verloop van een (beperkte) hoeveelheid komt het eerste betaalde abonnement in zicht. Voorbeeld: Dropbox is gratis tot 4 GB, daarna ga je ook voor Dropbox betalen!