

Content heeft een belangrijke rol in de informatievoorziening van organisaties. Bovendien groeit de hoeveelheid content door de huidige digitalisering enorm. Daarom is aandacht voor de plaats van content in een organisatie geboden. Daarnaast moeten we in staat zijn content gedurende de gehele levenscyclus adequaat te ondersteunen. Architectuur kan daarvoor een waardevol instrument zijn. Dit artikel beschrijft een referentie-architectuur voor ECM die als basis voor een stappenplan bruikbaar is.

Danny Greefhorst



Pragmatisch hulpmiddel

Een referentie-architectuur voor ECM

Content is een algemeen woord: het is de inhoud of het eindresultaat van een dienst of product (bron: Wikipedia). Enterprise content management (ECM) is een geformaliseerde manier van omgaan met content. Het omvat zowel de strategie als de methoden en tools die nodig zijn voor omgaan met content in processen van organisaties. ECM omvat de gehele levenscyclus van content en ook alle vormen waarin deze content zich kan manifesteren, zowel fysiek (papier) als digitaal.

In dit artikel bedoelen we met content 'ongestructureerde gegevens'. Dit zijn gegevens die geen voorgedefinieerd gegevensmodel hebben of niet georganiseerd zijn in een voorgedefinieerde structuur. Content bestaat vaak uit tekst, maar kan ook andere vormen hebben zoals afbeeldingen, geluid

of video. Bekende voorbeelden zijn webcontent (webpagina's), documenten, e-mail en berichten op social media.

In de praktijk is het lastig om een hard onderscheid te maken tussen gestructureerde en ongestructureerde gegevens. Vaak is er sprake van een combinatie tussen gestructureerd en ongestructureerd. Zo bevat een document vaak wel enige vorm van structuur, kan het ook gestructureerde gegevens bevatten (bijvoorbeeld in tabelvorm) en is de metadata van content ook gestructureerd. Anderzijds kunnen er bijvoorbeeld in databases (gestructureerde gegevens) ook teksten zijn opgenomen.

Er wordt ook wel gesproken over 'semi-gestructureerde gegevens'. Dat zijn gegevens waar wel enige vorm van structuur in zit. Denk bijvoorbeeld aan teksten die voorzien zijn van 'tags' (labels,

termen) die aangeven wat delen van teksten betekenen. Semi-gestructureerde gegevens maken onderdeel uit van ECM.

Architectuurprincipes

In dit artikel worden drie architectuurprincipes voor het omgaan met content beschreven. Het volgen van deze principes helpt bij het bewuster omgaan met content en zorgt dat meer waarde uit content wordt gehaald.

De gehele levenscyclus van content ondersteunen

Content heeft een levenscyclus van ontstaan tot vernietiging. De definitie van ECM geeft aan dat de gehele levenscyclus van content van belang is. De achterliggende gedachte is dat elke levensfase aandacht vraagt, goed ondersteund moet worden

en onderling goed aansluit. Professioneel beheer van content vloeit voort uit wet- en regelgeving. Het is daarom belangrijk om voor alle typen content binnen een organisatie na te denken over een aantal vragen:

- Wie is verantwoordelijk voor welke gegevens?
- Hoe worden gegevens bewerkt?
- Waar wordt data opgeslagen?
- Hoe wordt dit alles met IT ondersteund?

Dat begint met het benoemen van een eigenaar per contenttype. De eigenaar is verantwoordelijk voor het beleid rond content, inclusief het afbaken van de levenscyclus. Je zou dit 'governance' voor content kunnen noemen. Benoem welke applicaties worden gebruikt voor het beheren van bepaalde content en zorg dat deze het proces ondersteunen. Zorg voor bewustwording van de verschillen in karakteristieken van verschillende



typen content en de levensfasen waarin de content verkeert.

Content voorzien van relevante meta-data

Metadata zijn gegevens die beschrijven wat content is en waarop deze is gebaseerd. Zonder metadata als titel, auteur, versienummer en datum is content minder waardevol. Metadata bepalen de toegankelijkheid van de content, maar zijn ook essentieel voor archivering. Metadata kan zijn opgenomen in de content (embedded metadata) of apart van de content worden opgeslagen. Via metadata kunnen, door het gebruik van gemeenschappelijke termen, gegevens gekoppeld worden, zowel gestructureerde als ongestructureerde gegevens. Daarbij wordt bij voorkeur maximaal gebruik gemaakt van bestaande vocabulaires, thesauri en ontologieën zoals Dublin Core. Door

het is ontstaan. Het expliciet toegankelijk maken van deze content zorgt voor maximaal hergebruik. Meta-datering alleen is niet voldoende om toegankelijkheid te borgen. Belangrijk is dat content vindbaar is. Daar horen duidelijke afspraken bij over waar bepaalde content te vinden is. Dat kan bijvoorbeeld in de vorm van een portaal, waarbij de navigatiestructuur en gebruikersvriendelijkheid expliciet aandacht vragen. Er moet ook worden nagedacht over autorisaties op content. Uitgangspunt is bijvoorbeeld dat content die geen gevoelige gegevens bevat voor iedereen toegankelijk is. Idealiter zou dit via één zoekingang vindbaar moeten zijn. Denk bij het maken van content ook expliciet na over hoe de content door anderen afgenomen en gebruikt zal worden. Content wordt toegankelijker naarmate het in een meer gestructureerde kennisbank wordt beheerd en is voorzien van tags die betekenis geven.

Referentiemodel

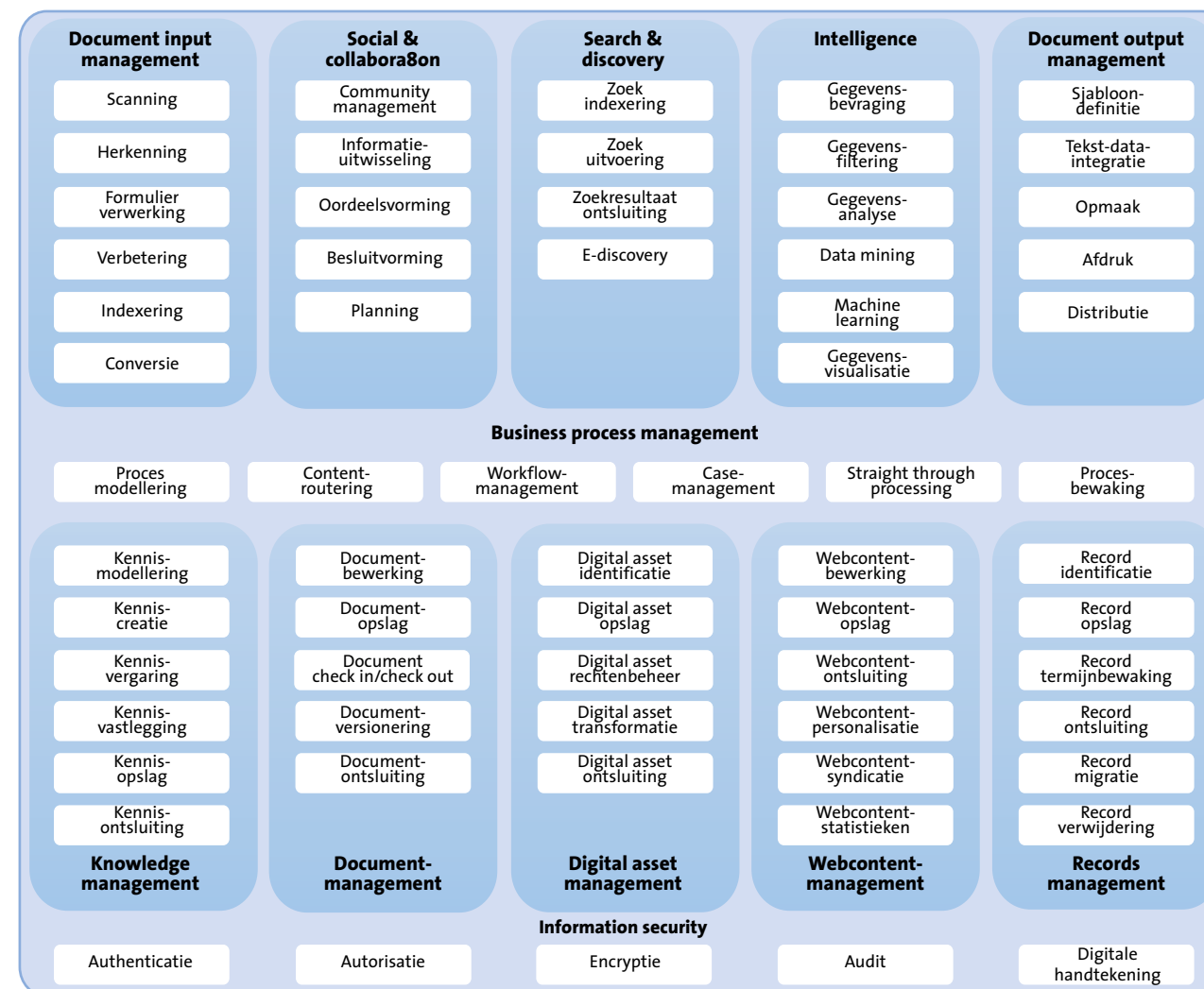
De kern van de referentie-architectuur is een referentiemodel waarin de belangrijkste gebieden en functies op het gebied van ECM zijn beschreven (figuur 1). Het referentiemodel is een concretisering van de bovengenoemde architectuurprincipes en algemene kennis over het onderwerp. Goede afbakening en kiezen van het juiste abstractieniveau blijft lastig. Voor de afbakening is aangesloten op de gebieden die in het algemeen worden gezien als onderdeel van ECM (zie bijvoorbeeld: AIIM^[1]). Daarnaast is technische infrastructuur buiten de scope gelaten. De benoemde gebieden zijn: document input management, social & collaboration, search & discovery, intelligence, output management, business process management, knowledge management, documentmanagement, digital asset management en information security. **Figuur 2** geeft van elk gebied een korte omschrijving. Binnen de verschillende gebieden is een verdieping gemaakt naar functies. Dit zijn functies die worden ondersteund door geautomatiseerde informatievoorziening. Een aantal van deze functies zijn meer applicatief, anderen meer infrastructureel. Het referentiemodel is daarbij onderdeel van de informatiearchitectuur en biedt ondersteuning aan allerlei processen. Voor het identificeren van de functies is vooral vanuit procesperspectief gekeken; hoe kunnen bepaalde processen en activiteiten ondersteund worden?

gegevens expliciet aan elkaar te verbinden krijgen ze meer betekenis en kunnen ze ook snel worden gevonden in andere contexten. Het is daarom belangrijk na te denken over de vraag welke metadata gewenst is voor verschillende typen content. Deze meta-datering moet expliciet worden ondersteund in applicaties waarin content wordt gecreëerd en vastgelegd, waarbij de gebruiker bij voorkeur zo min mogelijk wordt lastiggevalen. Ervaring leert dat expliciet vragen om meer metadata tot hoge drempels bij gebruikers leidt. Een deel van de metadata kan ook automatisch worden afgeleid. Denk bijvoorbeeld aan het afleiden van de naam van een auteur op basis van zijn inlognaam en een datum op basis van het huidige tijdstip. Meta-datering dient te worden geïntegreerd in het proces.

Relevante content expliciet toegankelijk maken

Veel content is breder relevant dan de plaats waar

Meta-datering moet expliciet worden ondersteund in applicaties waarin content wordt gecreëerd en vastgelegd



Figuur 1. Referentiemodel

Gebied	Omschrijving
Document input management	Het ontvangen en verwerken van documenten.
Social & collaboration	Het in contact staan met anderen en gezamenlijk bijdragen aan doelstellingen.
Business process management	Het verbeteren van de prestatie van de organisatie door het beheersen en optimaliseren van bedrijfsprocessen.
Search & discovery	Het zoeken en ontdekken van relevante gegevens.
Intelligence	Het halen van informatie uit gegevens.
Document output management	Het verspreiden van documenten aan anderen.
Knowledge management	Het beheersen van de levenscyclus van kennis.
Documentmanagement	Het beheersen van de levenscyclus van documenten.
Digital asset management	Het beheersen van de levenscyclus van digital assets.
Webcontentmanagement	Het beheersen van de levenscyclus van webcontent.
Records management	Het beheersen van de levenscyclus van records.
Information security	Het borgen van de beschikbaarheid, integriteit en vertrouwelijkheid van gegevens.

Figuur 2. Omschrijving van de gebieden in het referentiemodel

[1] Association for Information and Image Management: What is Enterprise Content Management (ECM)? www.aiim.org/What-is-ECM-Enterprise-Content-Management.aspx



Voor de afbakening van de individuele functies is ook gekeken naar een logische clustering van functionaliteit. Functies zijn gescheiden als ze door verschillende IT-oplossingen kunnen worden ondersteund. Meta-datering is niet als losse functie opgenomen in het model omdat het een integraal onderdeel is van het herkennen/identificeren en bewerken van content.

Veel ECM-leveranciers hebben zich gericht op het leveren van generieke ECM-oplossingen die met alle vormen van content overweg kunnen. In de praktijk hebben bepaalde typen content toch verschillende karakteristieken, verschillende levenscycli, en vragen zij om verschillende vormen van IT-ondersteuning. Daarom is er in het

Gegevens zijn steeds meer 'digital born' waardoor scannen en printen steeds minder relevant worden

referentiemodel voor gekozen om content meer gedifferentieerd uit te werken. Het model maakt onderscheid tussen:

- Semantische content: gegevens die zijn uitgedrukt in termen van een ontologie (kennismodel).
- Sociale content: gegevens die worden uitgewisseld tussen mensen. Dit zijn berichten die worden uitgewisseld op social media, maar ook bijvoorbeeld e-mail.
- Document: een bundeling van verschillende vormen van content die als geheel bewerkt kan

worden en in bestandsvorm wordt opgeslagen.

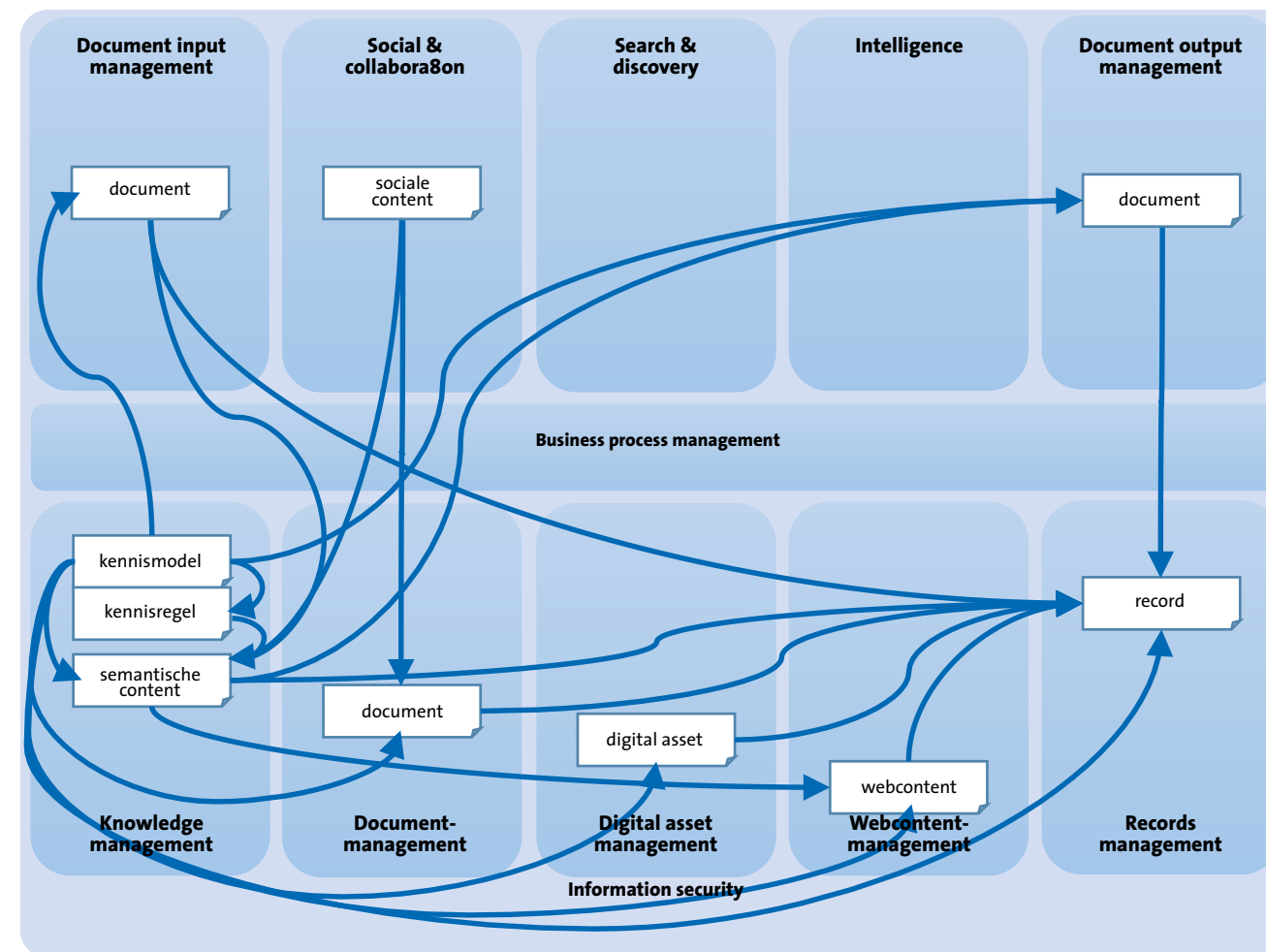
- Record: een gegeven dat bewust bewaard wordt als bewijs van een actie.
- Webcontent: een bundeling van verschillende vormen van content bedoeld voor presentatie in een webbrowser.
- Digital asset: content die een bepaalde waarde vertegenwoordigt en typisch vergezeld gaat van een bepaald recht van gebruik. In veel gevallen is dat multimediale content zoals afbeeldingen, geluid of video.

Merk op dat dit overlappende contenttypen zijn: content kan bijvoorbeeld zowel semantische content als webcontent zijn. Deze combinatie is de basis van het semantische web; door webpagina's te voorzien van betekenis zijn ze beter te begrijpen en beter bruikbaar. Zo kan bijvoorbeeld een adres op een webpagina herkend worden, wat het makkelijk maakt het aan een adresboek toe te voegen. De overlap in typen content geeft aan dat content verschillende rollen kan hebben.

In het model is relatief veel aandacht voor 'documenten'. Dit omdat documenten altijd een belangrijke manifestatie van content waren. In de toekomst zullen documenten minder belangrijk worden. Gegevens zijn steeds meer 'digital born' waardoor document input management en document output management, ofwel scannen en printen, steeds minder relevant worden. Los daarvan is een document een relatief willekeurige groepering van gegevens wat het vinden en hergebruik bemoeilijkt. Knowledge management wordt steeds belangrijker. Denk met name aan semantische webtechnologie, linked data en het geautomatiseerd redeneren op basis van kennisregels. Gartner noemt dit de 'algorithm economy'. **Figuur 3** toont welk type content in de verschillende gebieden wordt gecreëerd en wat mogelijke overgangen zijn. Een gescand document wordt typisch een record in records management. Binnen knowledge management zijn ook kennismodel en kennisregels weergegeven. Deze liggen ten grondslag aan de semantische content. Het kennismodel zou bij meta-datering ook in andere typen content gebruikt moeten worden om de toegankelijkheid te verhogen. Content wordt daarbij gekoppeld aan relevante termen uit het kennismodel (ontologie of thesaurus).

Gebruik referentie-architectuur

De referentie-architectuur kan op verschillen-



Figuur 3. Relatie tussen gebieden en contenttypen

de manieren worden gebruikt. Het kan gebruikt worden om te komen tot een roadmap van veranderingen. De architectuurprincipes kunnen worden vergeleken met de huidige praktijk. Dat biedt inzicht in mogelijke verbeteringen. Daarbij kan een scoring over de wenselijkheid van verbeteringen behulpzaam zijn. Een handig instrument daarbij is bijvoorbeeld 'MoSCoW' met scoringswaarden als 'Must have', 'Should have', 'Could have' en 'Would be nice to have'. De referentie-architectuur kan ook worden gebruikt als basis voor een organisatie-specifieke architectuur. De architectuurprincipes en het referentiemodel kunnen dan worden aangepast en aangevuld voor de specifieke organisatiecontext. De termen kunnen dan ook meer in lijn worden gebracht met termen die gebruikelijk zijn in de organisatie. Het referentiemodel heeft daarnaast ook intrinsieke waarde; het is een neutraal kader voor de organisatieontwikkeling.

Conclusie

Dit artikel schetste een referentie-architectuur voor ECM, bestaande uit een aantal architectuurprincipes en een referentiemodel. Het is vooral een pragmatisch hulpmiddel dat inzicht biedt op allerlei aspecten van ECM. Het kan helpen bij het inrichten van ECM. Het model wordt doorontwikkeld, waarbij feedback van anderen welkom blijft. Kennis delen is per slot van rekening een essentieel onderdeel van onze kenniseconomie.

Danny Greefhorst (dgreefhorst@archixl.nl) is directeur en principal consultant bij ArchiXL en werkzaam als enterprise- en informatiearchitect. Hij is actief in het architectuurvakgebied en publiceert hier ook regelmatig over. Hij is voorzitter van Stichting Digital Architecture en voorzitter van de special interest group Architectuur van het Ngi-NGN.