



publitas

SCRIPTIE

DATA CONSOLIDATIE
WEPUBLISH EN PUBLITAS

7 april 2024



Auteur: Nick Reintjes
Versie: 1.1

Colofon

ORGANISATIE

Naam: Publitas / WePublish B.V.
 Bezoekadres: Kleibultweg 35
 Postcode: 7575 BM
 Plaats: Oldenzaal
 Telefoon: 0541-571680

BEDRIJFSBEGELEIDING

Naam: Marc Lohuis
 Functie: Technical Director
 Telefoon: 06 22 08 17 14
 E-mail: marc@publitas.com

ONDERWIJSINSTELLING

Naam: Saxion University of Applied Sciences
 Bezoekadres: Handelskade 75
 Postcode: 7417 DH
 Plaats: Deventer
 Telefoon: 088 019 8888

AFSTUDEERBEGELEIDER

Afstudeerbegeleider: Hans Schokker
 E-mail: j.a.schokker@saxion.nl
 Telefoon: 06 13 63 01 43

STUDENT

Naam: Nick Reintjes
 Studentnummer: 452339
 E-mail: n.reintjes@publitas.com
452339@student.saxion.nl
 Telefoon: 06 38 62 24 32
 Opleiding: HBO-ICT Business

Voorwoord

Voor u ligt de scriptie "Dataconsolidatie WePublish en Publitas". Deze scriptie is geschreven voor de laatste fase van mijn studie ICT-Business aan het Saxion. Sinds november 2023 ben ik bezig geweest met het uitvoeren van het onderzoek en het schrijven van de scriptie. Deze scriptie geeft inzicht in hoe WePublish en Publitas de datavoorziening op strategisch en architectuurniveau effectief kunnen consolideren.

Ik wil mijn bedrijfsbegeleider Marc Lohuis bedanken voor de kans om mijn onderzoek uit te voeren binnen de organisatie. Tijdens het uitvoeren van mijn scriptie hebben we de afspraak gemaakt dat ik zelf de regie in handen had en altijd kon aankloppen voor advies of begeleiding. Daarnaast wil ik mijn begeleider vanuit het Saxion, Hans Schokker, hartelijk bedanken voor de begeleiding en goede en concrete feedback tijdens het uitvoeren van het afstudeerproject.

Ik wens u veel leesplezier toe.

Nick Reintjes

Oldenzaal, 07 april 2024.

Samenvatting

In november 2022 heeft Publitas de organisatie WePublish overgenomen. Beide organisaties zijn actief in de digitale folder branche en houden zich voornamelijk bezig met het digitaliseren daarvan. Momenteel ligt er een uitdaging voor Publitas omdat er geen universele tooling of IT-architectuur is. Het hoofddoel van dit onderzoek is daarom om een advies en ontwerp te presenteren voor de toekomstige IT-architectuur die is afgestemd op de geconsolideerde interne en externe folderdata. Op basis van het hoofddoel is de volgende hoofdvraag ontstaan: "Hoe kan Publitas, na de overname van WePublish, op een effectieve manier de data-toepassingen van beide partijen consolideren en wegzetten in een datawarehouse-omgeving, om dit vervolgens te visualiseren?".

Om antwoord te geven op de hoofdvraag, is de informatie/datastrategie van beide organisaties onderzocht met betrekking tot folderdata en is beoordeeld of deze strategie past bij de nieuwe situatie. Voor dit doel is een advies geformuleerd. Daarnaast is er aandacht besteed aan het in kaart brengen van de huidige situatie op het gebied van IT-architectuur en dataniveau, waarbij ook naar de gemeten interne en externe folderdata is gekeken. Vervolgens is geanalyseerd hoe deze situatie aangepast moet worden om tot de gewenste IT-architectuur te komen en hiervoor is een ontwerp opgenomen. Tot slot zijn er wireframes opgesteld voor de gewenste situatie.

Wat betreft de strategie omtrent informatievoorziening komt een tweedelig advies naar voren. Publitas moet zich richten op het eigen platform van de retailers, terwijl WePublish zich breder moet oriënteren, met name richting affiliate/folderplatformen. Verder is de huidige IT-architectuur en de data-tracking in kaart gebracht. Publitas heeft een IT-architectuur omgeving binnen AWS, terwijl WePublish binnen Azure opereert. De datakwaliteit en schaalbaarheid zijn goed ingeregeld bij Publitas, maar bij WePublish is dit niet het geval. De getrackte events binnen de folder viewers verschillen minimaal van elkaar.

Als er gekeken wordt naar de toekomstige situatie moet de IT-architectuur geconsolideerd worden binnen de AWS-omgeving van Publitas, mede doordat de organisatie heeft besloten dat de WePublish retailers gemigreerd moeten worden naar de Publitas tooling. Verder moeten de API's worden ingeregeld middels Apache Airflow-pipelines. DevOps zal daarnaast verantwoordelijk worden voor het volledige beheer van de geconsolideerde IT-architectuur. Tot slot zijn er wireframes opgesteld in de Publitas huisstijl die Publitas moeten helpen bij het visualiseren van de toekomstige situatie.

Op basis van de deelconclusies kan de hoofdvraag beantwoord worden. Publitas dient zich te blijven richten op de eigen platformen, terwijl WePublish zich moet concentreren op de folderplatformen/affiliate platformen. De IT-architecturen van beide organisaties dienen te worden geconsolideerd binnen de AWS-omgeving, waar ze interne en externe data samenbrengen. Deze data kan vervolgens effectief worden gevisualiseerd aan de hand van de opgestelde wireframes, rekening houdend met de belangen van de primaire stakeholders, de gestelde requirements en de literatuur van Stephan Few.

Op basis van de enquêteresultaten van de customer success managers blijkt dat zij nog te weinig kennis hebben en de mogelijkheden niet volledig benutten als het gaat om het gebruik van folderdata. Op basis hiervan geldt dus de aanbeveling om de customer succes managers verder te helpen door hen de waarde van data in te laten zien, hier speelt de kennis en expertise van WePublish een belangrijke rol. Tot slot is het onderzoek en de vormgeving van de wireframes intern uitgevoerd met de primaire stakeholders. Hierbij geldt de aanbeveling om ook de externe stakeholders, dus de retailers erbij te betrekken.

Inhoudsopgave

Voorwoord	3
Samenvatting	4
Figuren- en tabellenlijst	6
1 Inleiding	8
1.1 Kernactiviteiten WePublish.....	8
1.2 Kernactiviteiten Publitas	9
1.3 Probleemstelling en doelstellingen.....	9
1.4 Hoofd- en deelvragen	10
2 Theoretisch kader	11
2.1 Kernbegrippen	11
2.2 Theorieën, modellen en onderzoeksmethoden	13
3 Methodologie	16
3.1 Data verzameling	16
3.2 Methodiek per deelvraag.....	16
3.3 Onderzoeksmodel	22
4 Resultaten	23
4.1 Deelvraag 1	23
4.2 Deelvraag 2	25
4.3 Deelvraag 3	28
4.4 Deelvraag 4	30
4.5 Deelvraag 5	37
5 Conclusie	45
6 Discussie	46
7 Aanbevelingen	47
Bibliografie	48
Bijlage 1 – Interview Marc Lohuis	50
Bijlage 2 – Interview Paul Raya	53
Bijlage 3 – 5 Krachten model van porter	55
Bijlage 4 – SWOT-Analyse	58
Bijlage 5 – Interview Toby Nijboer	61
Bijlage 6 – Interview Umer Khan.....	64
Bijlage 7 – IT-architectuur WePublish.....	67
Bijlage 8 – Data-mapping WePublish & Publitas	70

Bijlage 9 – Stakeholder analyse	87
Bijlage 10 – Interview Ibe Hundling	90
Bijlage 11 – Interview Ramon Eerdman	92
Bijlage 12 – Gap-Analyse van Ist naar Soll	95

Figuren- en tabellenlijst

Figuur 1 Gap analyse	14
Figuur 2 Voorbeeld niet-opvallende en opvallende kleuren	15
Figuur 3 Voorbeeld KPI button.....	15
Figuur 4 Waar gaat de aandacht naartoe	15
Figuur 5 Onderzoeksmodel	22
Figuur 6 Grafiek GFK	23
Figuur 7 SWOT-analyse	24
Figuur 8 WePublish IT-architectuur	25
Figuur 9 Publitas data opslag	26
Figuur 10 Publitas data tracking	26
Figuur 11 Publitas IT-architectuur.....	27
Figuur 12 Gewenste IT-architectuur	31
Figuur 13 IT-architectuur Snowplow & Databricks	32
Figuur 14 IT-architectuur API Data & MsSQL.....	33
Figuur 15 PowerBI folder filters	34
Figuur 16 Feitentabel oud en nieuw	34
Figuur 17 STAR-model.....	35
Figuur 18 Wireframe KPI's & kanalen	41
Figuur 19 Wireframe folder inzichten.....	42
Figuur 20 Wireframe kanalen en source.....	42
Figuur 21 Wireframe jaar vergelijking	43
Figuur 22 Wireframe folder overzicht	43
Figuur 23 5 Krachten model porter.....	55
Figuur 24 SWOT-analyse	58
Figuur 25 WP IT-architectuur folder viewer en tracking.....	67
Figuur 26 WePublish IT-architectuur Cosmos.....	67
Figuur 27 WePublish IT-architectuur MySQL.....	67
Figuur 28 Data API's	67

Figuur 29 WePublish IT-architectuur Azure data pipelines	68
Figuur 30 WePublish IT-architectuur MsSQL database	68
Figuur 31 WePublish IT-architectuur dashboarding	68
Figuur 32 WePublish IT-architectuur	69
Figuur 33 WePublish events (opgedeeld per event).....	77
Figuur 34 WePublish dimensie tabellen	78
Figuur 35 View total fact table.....	79
Figuur 36 Tabel totalfact.....	79
Figuur 37 Databricks unique views	85
Figuur 38 Databricks page views.....	85
Figuur 39 Databricks page events	85
Figuur 40 Dimensie tabellen Databricks	86
Figuur 41 Categorisatie Stakeholders	88
Figuur 42 Prioritering stakeholders	89
Figuur 43 Gap analyse.....	95
Tabel 1 Kernbegrippen.....	12
Tabel 2 Beschrijving feitentabel.....	35
Tabel 3 Functionele requirements.....	38
Tabel 4 niet functionele requirements	39
Tabel 5 MoSCoW priotisering	40
Tabel 6 Data tracking WePublish (opgedeeld per event)	70
Tabel 7 Snowplow data tracking (opgedeeld per event)	80

1 Inleiding

In november 2022 heeft Publitas de organisatie WePublish overgenomen. Sindsdien zijn er veel inspanningen geleverd om de organisaties dichter bij elkaar te brengen en elkaars werkwijze te begrijpen.

WePublish is een softwareontwikkelingsbedrijf dat oorspronkelijk werd opgericht onder de naam Online Publisher in 2006. Het bedrijf richt zich voornamelijk op het digitaliseren van papieren advertentie folders en magazines en het is gevestigd in Oldenzaal. WePublish heeft ongeveer 10-15 Fte's in dienst.

Publitas is, net als WePublish, een softwarebedrijf en werd ook opgericht in 2006. Publitas richt zich eveneens op het digitaliseren van papieren advertentie folders en magazines. Echter gaat Publitas breder te werk en is internationaal actief in veel verschillende landen. Publitas heeft ongeveer 80 Fte's in dienst verspreid over de hele wereld. Opvallend is dat Publitas geen fysiek kantoor heeft en als een volledig remote bedrijf opereert.

1.1 Kernactiviteiten WePublish

WePublish is, zoals beschreven in het voorgaande hoofdstuk, een softwareontwikkelingsbedrijf dat gespecialiseerd is in het digitaliseren van advertentiefolders en magazines. Deze specialisatie kan echter beknopt lijken. WePublish biedt retailers een tool aan, genaamd het Folder Management Platform (FMP), waarmee zij een PDF van een folder of magazine kunnen verwerken.

Met het FMP kunnen retailers de PDF van de folder uploaden en interacties toevoegen in de vorm van hyperlinks naar producten op de website, maar ook naar bijvoorbeeld een YouTube-video of een interne inhoudsopgave in de online folder. Vervolgens kunnen ze de verwerkte folder digitaal tonen op een zelfgekozen locatie op hun eigen platform/website. Hiervoor moeten ze de bijbehorende viewer van WePublish gebruiken en integreren op de website, de viewer is een tool die het mogelijk maakt om de folder te tonen op een gekozen locatie. Als een retailer ervoor kiest om de folder zelf te verwerken, wordt dit een selfservice klant genoemd. WePublish biedt echter ook een full-service optie aan, waarbij de retailer de folder met bijbehorende interacties aanlevert en een servicemedewerker de verwerking uitvoert.

Naast de tooling voor de eigen platformen/websites van retailers heeft WePublish ook het folderplatform Folders.nl in haar beheer. Een folderplatform is in het kort een verzamelwebsite voor digitale folders, waarop een groot aanbod aan advertentiefolders van verschillende retailers te vinden is. Dit platform draait ook op back-end en viewer van WePublish.

Binnen de door WePublish aangeboden viewer is ook een ingebouwde data tracker van Matomo opgenomen. WePublish meet aan de hand van deze tracker het gedrag van de consument, waarbij rekening wordt gehouden met de AVG-gegevens wetgeving. De verzamelde gegevens worden vervolgens opgehaald, verwerkt en geanalyseerd. Als een retailer betaalt voor het dashboard, krijgen zij ook zelf inzicht in de prestaties.

Naast het dashboard biedt WePublish ook consultancy- en optimalisatie trajecten aan. Tijdens deze trajecten wordt aan de hand van data gekeken hoe een retailer meer uit de digitale folder kan halen en zo een hogere conversie kan behalen.

1.2 Kernactiviteiten Publitas

Publitas is, zoals beschreven, net als WePublish een softwareontwikkelingsbedrijf dat zich richt op het digitaliseren van papieren advertentiefolders en magazines. Publitas biedt hiervoor, net als WePublish, een eigen tool aan die dit mogelijk maakt, genaamd 'Publitas Play'.

Het productieproces rondom het uploaden, verwerken en live zetten van online folders werkt in grote lijnen hetzelfde als dat van WePublish. Het verschil ligt echter in de softwarecatalogus van Publitas. Naast de tooling biedt Publitas meerdere softwareoplossingen aan. Enkele voorbeelden van softwaretoepassingen zijn een winkelmand interactie, waarbij een retailer producten direct kan laten toevoegen aan een online winkelmand op hun eigen website. Via WhatsApp consumentenproducten kunnen toevoegen aan hun winkelmand die ze hebben gezien via de digitale folder. Hetzelfde geldt voor e-mail en er zijn nog veel meer toepassingen beschikbaar.

Kortom, Publitas heeft een uitgebreid assortiment aan toepassingen die gerelateerd zijn aan de digitale folder. Wat betreft datavoorziening schiet Publitas echter tekort in vergelijking met WePublish. Hierbij doel ik met name op de data-consultancy van Publitas. Hoewel data wordt getrackt via de Snowplow-data tracker, wordt hier nog onvoldoende mee gedaan naar mijn mening. Publitas heeft namelijk een beperkt dashboard en biedt (nog) geen optimalisatie- of consultancy trajecten aan.

1.3 Probleemstelling en doelstellingen

Nu de organisaties WePublish en Publitas zijn samengegaan onder de naam Publitas, dient er een grondige analyse te worden uitgevoerd met betrekking tot verschillende vraagstukken. Eén van de voornaamste vraagstukken betreft de strategische koers van de organisatie op het gebied van de data- en informatievoorziening gericht op folderdata.

Zowel Publitas als WePublish beschikken over een eigen IT-architectuur zoals hierboven beschreven, wat betekent dat data op meerdere plaatsen is opgeslagen. Het is noodzakelijk om zowel de interne als externe data te analyseren om vast te stellen hoe deze op de beste wijze kunnen worden samengebracht binnen een geconsolideerde IT-architectuur.

Indien deze data- en IT-architectuur niet wordt geconsolideerd binnen een geschikte datawarehouse oplossing, kan dit mogelijk resulteren in een inefficiënte datavoorziening. Het is voordelig om de data- en IT-architectuur te consolideren om bijvoorbeeld de volgende redenen: centrale opslag zorgt voor efficiëntie bij het beheren en gebruiken van data. Beheerders en gebruikers hebben een enkel punt van toegang, wat het systeem vereenvoudigt. Verder, door het te consolideren op één locatie, is het beter te beveiligen doordat je dit kunt concentreren op één plek en tot slot kostenbesparing. Je maakt minder kosten doordat je maar één systeem hoeft te hosten.

Het project heeft dus meerdere doelstellingen. Om te beginnen is het belangrijk om de huidige strategische koers met betrekking tot de data- en informatievoorziening, gericht op folderdata, te analyseren. Deze analyse is essentieel om een goed toekomstbestendig ontwerp voor de IT-architectuur te realiseren verderop in het onderzoek. Op basis van bovenstaande analyse wordt een passend advies opgesteld over of de huidige koers de juiste koers is, of dat deze moet worden bijgesteld.

Het tweede doel is om een grondige analyse uit te voeren van de twee IT-architecturen en deze zorgvuldig te modelleren. Om verderop in het onderzoek een goed ontwerp en advies op te stellen voor de toekomstige IT-architectuur, is het namelijk van belang dat er eerst naar de huidige situatie wordt gekeken. Hierbij worden de zwakke en sterke punten geïdentificeerd en wordt bepaald waar

wel en niet op voortgeborduurd kan worden. Er wordt gekeken naar het beheer van de architecturen, schaalbaarheid, beveiliging en datakwaliteit. Daarnaast wordt er aandacht besteed aan welke data wordt bijgehouden, de verschillen daarin en hoe deze data effectief kan worden samengevoegd. Het is van essentieel belang dat de klant geen waardeverlies zal ervaren in de toekomst als gevolg van de fusie van beide IT-architecturen.

Het derde doel is dat de geanalyseerde data op een correcte manier wordt opgenomen binnen een geconsolideerde IT-architectuur. Het doel hiervan is om een goed doordacht ontwerp op te zetten dat Publitas in de toekomst kan implementeren. Ten slotte is onderzocht hoe een dashboard eruit zou moeten zien in lijn met de nieuwe situatie na de fusie, rekening houdend met de primaire stakeholders.

Op basis van het voorgaande is het hoofddoel als volgt geformuleerd: De doelstelling van het afstudeerproject is om, na de overname van WePublish door Publitas, een advies en ontwerp te presenteren voor een toekomstige IT-architectuur die is afgestemd op de geconsolideerde folderdata van beide organisaties.

1.4 Hoofd- en deelvragen

De onderzoeksvraag van dit project luidt als volgt:

“Hoe kan Publitas, na de overname van WePublish, op een effectieve manier de data toepassingen van beide partijen consolideren en wegzetten in een datawarehouse omgeving, om dit vervolgens te visualiseren.”

Om tot een goed en volledig antwoord te komen op de onderzoeksvraag zijn de volgende deelvragen opgesteld en uitgewerkt:

1. *Wat zijn de consequenties van de fusie voor de informatievoorziening strategie met betrekking tot folderdata van Publitas en welk advies is er voor de toekomstige strategie omtrent de informatievoorziening?*
2. *Hoe ziet de huidige data-architectuur eruit bij zowel WePublish als Publitas en wat zijn de sterke punten en zwakke punten ervan?*
3. *Hoe dient de geconsolideerde data van WePublish en Publitas vormgegeven te worden?*
4. *Hoe dient de toekomstige IT-architectuur eruit te zien voor Publitas, waarin zij de interne en externe folderdata kunnen opslaan?*
5. *Hoe kan de geconsolideerde data effectief worden gevisualiseerd om waardevolle inzichten te bieden aan de stakeholders?*

Leeswijzer

Hoofdstuk 2 geeft een overzicht van de kernbegrippen, de gebruikte theorieën, modellen en onderzoeksmethoden. In hoofdstuk 3 wordt per deelvraag de methodologie toegelicht. Vervolgens komen in hoofdstuk 4 de deelresultaten naar voren en op basis hiervan wordt in hoofdstuk 5 de conclusie en beantwoording van de hoofdvraag toegelicht. Tot slot worden in hoofdstuk 6 de beperkingen van het onderzoek besproken en worden de aanbevelingen gepresenteerd in hoofdstuk 7.

2 Theoretisch kader

In dit hoofdstuk zullen de belangrijkste begrippen, theorieën en modellen worden beschreven die tijdens het onderzoek worden aangehaald.

2.1 Kernbegrippen

Folder management platform (FMP) 1 & 2	Het Folder Management Platform is een tool die ontwikkeld is door WePublish die retailers en het serviceteam gebruiken om folders te verwerken. Onder "verwerken" verstaan we het uploaden van de pdf, het toevoegen van hyperlinks/interacties en het publiceren van de onlineversie op de gewenste locatie.
Publitas Play	Publitas play is een tool die ontwikkeld is door Publitas die retailers en het serviceteam gebruiken om folders te verwerken. Onder "verwerken" verstaan we het uploaden van de pdf, het toevoegen van hyperlinks/interacties en het publiceren van de onlineversie op de gewenste locatie. Binnen Publitas play is ook een in-app dashboard waar de retailer data kan inzien van de publicaties.
Folder viewer	Zowel WePublish als Publitas hebbe een folder viewer. Deze viewer moet door een retailer worden geïntegreerd op de locatie waar de digitale folder moet worden weergegeven. De viewer maakt dit mogelijk en bovendien is hieraan de datatracker gekoppeld die de gegevens bijhoudt.
Datawarehousing	Een datawarehouse is een centrale opslagplaats voor gestructureerde en semi-gestructureerde data. Binnen een datawarehouse kun je data ophalen uit verschillende bronnen, zoals databases, applicaties of andere toepassingen die data bevatten. Vervolgens wordt deze data vaak binnen het datawarehouse opgeschoond en gestandaardiseerd (Wat Is Een Datawarehouse?, n.d.). In een datawarehouse zijn verschillende elementen te vinden, waaronder een relationele database waarin de data wordt opgeslagen en beheerd, een ETL-oplossing (Extract, Transform, Load) en functionaliteiten voor analyses, rapportages en data mining (Oracle, n.d.).
Data lake	Een data lake is, net als een datawarehouse, een opslagplaats voor data. Echter kan binnen een data lake naast gestructureerde data en semi-gestructureerde data ook ongestructureerde data worden opgeslagen (Wat Is Een Data Lake?, n.d.) (Wat Is Een Data Lake? Data Lake Versus Warehouse, n.d.).
Datakwaliteit	Als het gaat over de datakwaliteit, gaat het over de geschiktheid van de data voor het uiteindelijke doel. De data dient dus betrouwbaar genoeg te zijn om hier bijvoorbeeld conclusies aan te verbinden. Naast de kwaliteit gaat datakwaliteit ook over het beheer van de data (ICTinformatiecentrum, 2023).
Data schaalbaarheid	Schaalbaarheid houdt in dat het IT-systeem zich goed kan aanpassen aan de behoefte en de werklast kan verwerken. In geval van een hoge werklast zal het moeten opschalen en bij een lage werklast zal het moeten downscalen. Het uitgangspunt is dat het systeem meeschaalt

met de werklust (Wat betekent schaalbaarheid voor systemen en services?, sd) (Azap, 2022).

API	Een API, wat staat voor Application Programming Interface, is een softwaretoepassing die het mogelijk maakt dat twee applicaties met elkaar kunnen communiceren. Bovendien maakt een API het uitwisselen van data tussen twee systemen mogelijk (Wat Is Een API En Wat Kan Je Ermee?, 2019).
S3 Loader & S3	Een Snowplow S3 Loader verwerkt gegevens uit een Amazon Kinesis-datastream en schrijft deze weg naar S3. S3 staat voor Simple Storage Service, een objectopslagservice. Een standaard opgebouwde Snowplow-pijplijn zal de S3 Loader op drie verschillende manieren gebruiken, namelijk: • De ruwe datastroom zorgt ervoor dat een back-up van de oorspronkelijke gegevens wordt behouden. • De verrijkte (enriched) data bestaat uit gegevens die zijn verrijkt met aanvullende informatie voordat ze worden opgeslagen in een S3-opslag. • De mislukte datastroom bevat gegevens die niet correct konden worden verwerkt of die niet voldoen aan de vereisten tijdens het verrijktingsproces. (S3 Loader Snowplow Documentation, 2023)
EC2 Cluster	Een EC2-cluster is een groep van virtuele machines die samenwerken om specifieke doelen te bereiken. In het geval van Publitas wordt het gebruikt voor dataprocessing, waarbij het data omzet naar een bruikbare vorm en het vervolgens doorstuurt. Dit is echter slechts een van de vele toepassingen. Bovendien zijn EC2-clusters ideaal om schaalbaarheid te garanderen via automatische schaling (Auto Scaling) en de belasting te verdelen door middel van elastische belasting balanceren (Elastic Load Balancing) (Cluster Management - Amazon Elastic Container Service, sd).
Folderplatform/affiliate platform	Een folderplatform is een platform/website waar meerdere folders op staan van verschillende retailers. Hier kan een consument aan de hand van de voorpagina op een folder klikken en vervolgens bekijken.

Tabel 1 Kernbegrippen

2.2 Theorieën, modellen en onderzoeksmethoden

In dit hoofdstuk worden de verschillende modellen en theorieën toegelicht.

2.2.1 5 Krachten model van Porter

Het vijfkrachtenmodel van Porter is een methodiek/ framework dat wordt gebruikt om de kracht en competitiviteit van concurrentie te beoordelen. Deze methode wordt ingezet om inzicht te krijgen in de concurrentie en een meer holistische blik te ontwikkelen voor het beantwoorden van de eerste deelvraag.

De vijf krachten:

1. De eerste kracht van het vijfkrachtenmodel van Porter behandelt de dreiging van nieuwe toetreders op de markt, oftewel hoe eenvoudig het is voor nieuwe concurrenten om de markt te betreden. Dit houdt in of er bijvoorbeeld aanzienlijke investeringen nodig zijn voor applicatiesoftware, of dat een concurrent zonder aanzienlijke kosten kan beginnen.
2. De tweede kracht van het vijfkrachtenmodel van Porter richt zich op de waarschijnlijkheid dat consumenten en voor Publitas in het bijzonder de retailers, kiezen voor een vervangend product. In dit scenario zou dit een andere viewer/verwerking tool betreffen.
3. De derde kracht van het vijfkrachtenmodel van Porter behandelt de onderhandelingskracht van de consument en in het geval van Publitas, de retailer. Het gaat erom of de aanbieder intensieve concurrentie ervaart, waardoor de markt minder aantrekkelijk wordt.
4. De vierde kracht van het vijfkrachtenmodel van Porter richt zich eveneens op de onderhandelingskracht, maar in dit geval betreft het de leveranciers. Voor Publitas, waar het om software gaat, wordt onderzocht of de leverancier een schaars product aanbiedt of dat er juist veel beschikbaar is op de markt.
5. De vijfde kracht van het vijfkrachtenmodel van Porter betreft de intensiteit van de concurrentie binnen de industrie. Hier zijn enkele vragen die je jezelf kunt stellen, zoals: hoeveel concurrenten zijn er, hoe groot zijn ze en is de industrie aan het groeien?

(Vijf Krachten Model Porter | Voorbeeld & Uitleg 5 Forces Model, n.d.)

2.2.2 SWOT-analyse

Een SWOT-analyse is een onderzoek dat de organisatie helpt bij het in kaart brengen van de sterke, zwaktes, kansen en bedreigingen. Door dit effectief te benutten, kun je strategische beslissingen rechtvaardigen en vaststellen welke aspecten van een organisatie extra aandacht verlangen of welke kansen er zich voordoen. Kansen en bedreigingen doen zich voor buiten de organisatie. Deze externe factoren kunnen zowel positieve als negatieve effecten hebben. Aan de andere kant spelen sterke en zwaktes zich binnen de organisatie af. Dit verwijst naar waar de organisatie goed in is en waarin zij minder sterk is (Gertjanschap, SWOT - Managementmodellensite, 2023). Deze methode is een van de methoden die bijdragen aan het beantwoorden van de eerste deelvraag.

2.2.3 Stakeholder analyse

De stakeholderanalyse heeft als doel het in kaart brengen van de belanghebbenden van het project en hun behoeften. Deze zogenoemde stakeholders worden vervolgens ingedeeld als primaire of secundaire stakeholders. Primaire stakeholders zijn sterk betrokken en secundaire minder. tijdens de uitvoering van de stakeholderanalyse loop je de volgende stappen door:

- 1) **Identificatie van stakeholders:** Hier breng je de belanghebbenden in kaart. Dus welke afdelingen, managers, medewerkers, enzovoort, zijn belanghebbend bij het project.
- 2) **Categorisering van stakeholders:** Hier deel je ze in als primair en secundair en wordt er gekeken of het interne, externe of interface stakeholders zijn. Interface stakeholders zijn niet direct betrokken bij het project, maar hebben wel een legitiem belang, zoals de maatschappij, universiteit of familie.
- 3) **Prioriteren van stakeholders:** Hier wordt naar twee aspecten gekeken, namelijk het belang van de stakeholder en de invloed/macht van de stakeholder. Je kunt hiervoor een matrix opstellen om te bepalen waar de prioriteiten moeten liggen.
- 4) **Bepalen van de stakeholdersaanpak:** De categorisering in de voorgaande stap is hiervoor een eerste stap. Dit onderdeel vormt de basis voor een goede stakeholderstrategie.

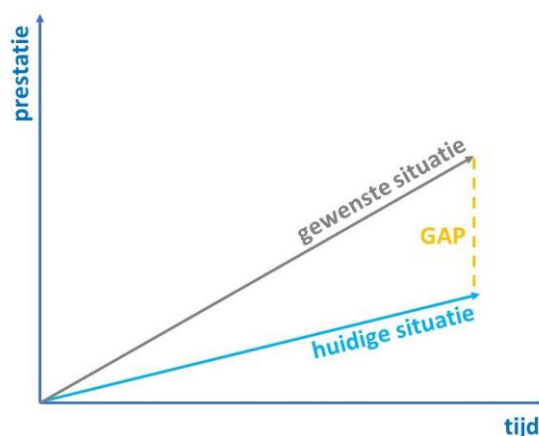
(Stakeholdersanalyse – projectmanagementsite, sd)

2.2.4 Data mapping

Data mapping is de eerste stap voor veel verschillende data-integratiedoeleinden. In de context van dit project is dit de eerste stap voor de dataconsolidatie. Het is namelijk van belang dat je weet welke data er beschikbaar is van beide databases en of deze data compatibel is. Data-mapping is in deze context het identificeren van overeenkomsten, redundante data of data die overbodig is. (Talend, sd)

2.2.5 GAP Analyse

Er wordt een GAP-analyse uitgevoerd om de huidige prestaties en de gewenste prestaties vast te leggen en vervolgens te bepalen welke aanpassingen nodig zijn om de gewenste prestaties te bereiken. Kort gezegd worden de huidige resultaten gemeten en vergeleken met de gewenste resultaten. Daarna wordt geanalyseerd wat er ontbreekt om de gewenste resultaten te behalen, wat het gevolg kan zijn van slecht opgezette strategieën, processen, ontbrekende vaardigheden, enzovoort (Gertjanschap, GAP-analyse vanuit Ist naar Soll - Managementmodellensite, 2023).



Figuur 1 Gap analyse

2.2.6 Literatuur Stephan Few (Dashboard Design)

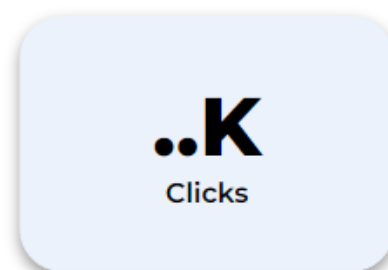
De wireframes die zijn opgesteld in de resultaten van deelvraag 5 zijn aan de hand van onder andere de interviews met de stakeholders opgesteld. Daarnaast is er gekeken naar de literatuur van Stephan Few, met name zijn boek 'Information Dashboard Design'. In dit hoofdstuk worden verschillende aspecten aangehaald waar rekening mee is gehouden tijdens het opstellen van de wireframes.

Het zien en begrijpen van data begint al voordat je je er bewust van bent. Het kan lastig zijn om details uit een grafiek of tabel te halen waar geen waarden zijn uitgelicht. Wanneer echter één of meerdere samenhangende aspecten oplichten, zullen deze punten eruit springen en opvallen (Few, Information Dashboard Design, 2006). Dit hangt samen met kleurgebruik. Er zijn standaardkleuren en benadrukkende kleuren, ook wel felle kleuren genoemd. Door deze goed samen te gebruiken, kun je de datapunten die van belang zijn benadrukken. In het opgestelde dashboard is hier rekening mee gehouden door een neutrale standaardkleur als achtergrond te gebruiken, de assen met donkere tekst weer te geven en de data waar het om draait te tonen met felle kleuren. Bij het kiezen van deze kleuren wordt ook rekening gehouden met de huisstijl van de platformen waarvan de data afkomstig is (Few, Information Dashboard Design, 2006).



Figuur 2 Voorbeeld niet-opvallende en opvallende kleuren

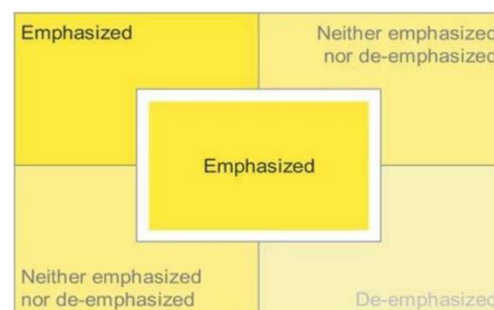
Verder geeft de literatuur aan dat je een grote hoeveelheid aan data kunt tonen door deze op te nemen in samenvattende datapunten. Denk hierbij bijvoorbeeld aan gesommeerde data of gemiddelden (Few, Information Dashboard Design, 2006). Zie figuur 3 voor een voorbeeld van een KPI-button. Hier zie je ook de neutrale achtergrond kleur met de benadrukkende data kleur.



Figuur 3 Voorbeeld KPI button

Om een goed en doelgericht dashboard op te stellen, dienen er zo min mogelijk niet-data-pixels te worden opgenomen in het dashboard en moeten de datapixels worden uitgelicht (Few, Information Dashboard Design, 2006)

Uit de literatuur blijkt dat de gebruiker het snelst kijkt naar het midden en de linkerbovenhoek. Dit heeft deels te maken met het feit dat we in westerse landen linksboven beginnen met lezen en schrijven. Daarom zijn de rechterbovenhoek en linkeronderhoek daarna het meest interessant en de rechteronderhoek het minst interessant (Few, Information Dashboard Design, 2006)). Bij de ontwikkeling van de wireframes is hier rekening mee gehouden door de belangrijkste informatie altijd in het midden te plaatsen, met de KPI-buttons aan de linkerkant.



Figuur 4 Waar gaat de aandacht naartoe

3 Methodologie

In dit hoofdstuk wordt eerst toegelicht hoe de data is verzameld, hoe de keuze is gemaakt welke data te verzamelen, het verloop van het onderzoek en hoe deze data wordt geanalyseerd. Vervolgens wordt per deelvraag gekeken naar de gebruikte methoden en de betrouwbaarheid en validiteit ervan.

Tot slot is het onderzoeksmodel toegevoegd, inclusief een beschrijving van hoe de verschillende deelvragen op elkaar aansluiten. Kortom, het biedt een uitleg van de rode draad van het onderzoek.

3.1 Data verzameling

Om tot een goede dataverzameling te komen, is per deelvraag gekeken naar de geschikte methoden, hiervoor is het DOT-Framework gebruikt. Deelvraag 1 tot en met 3 zijn beschrijvende deelvragen. Hier is gekeken naar de huidige staat van de informatie/data-strategie, IT-architectuur en de data-tracking. Tijdens de uitwerking is hier veel informatie ingewonnen via expertinterviews, data-analyse en eigen observaties en is dit onderbouwd met literatuur. Als je vervolgens naar deelvraag 4 en 5 kijkt, zijn dit ontwerp/advies vragen. Dus hoe gaan we van de oude situatie naar de gewenste toekomstige situatie.

3.2 Methodiek per deelvraag

In het hoofdstuk Methodiek wordt per deelvraag toegelicht welke data is verzameld, hoe deze data is verzameld en hoe de betrouwbaarheid en validiteit zijn gewaarborgd.

3.2.1 Deelvraag 1

Deelvraag 1 luidt als volgt: “Wat zijn de consequenties van de fusie voor de informatievoorziening strategie met betrekking tot folderdata van Publitas en welk advies is er voor de toekomstige strategie omtrent de informatievoorziening?”.

3.2.1.1 Dataverzameling

Semigestructureerde interviews

Tijdens het uitvoeren van deelvraag 1 zijn twee semigestructureerde interviews opgesteld en uitgevoerd met als doel de huidige strategie en missie en visie in kaart te brengen. Er is voor deze methode gekozen omdat de visie/missie niet duidelijk is beschreven als het gaat over folderdata. Tijdens de interviews zijn onderwerpen als de huidige missie en visie rondom folderdata behandeld, maar ook is hier input opgevraagd voor de SWOT-analyse.

In deelvraag 1 is de data en CRM-lead van Publitas en de technisch directeur van WePublish, geïnterviewd. Er is voor hen gekozen omdat zij beide verantwoordelijk zijn voor de data-afdeling en dus ook medeverantwoordelijk zijn voor de van de strategie rondom folderdata. Deze gesprekken zijn opgeschoond en in de bijlage opgenomen.

5 Krachten model van Porter & SWOT-Analyse

Om tot een toekomstig advies te komen met betrekking tot de data- en informatiestrategie is het 5-krachtenmodel Van Porter opgesteld om de kracht van de concurrentie in kaart te brengen. Vervolgens is er met de input van onder andere de semigestructureerde interviews en het 5-krachtenmodel van Porter een SWOT-analyse opgesteld. Binnen de SWOT-analyse is gekeken waar de sterktes, zwaktes, kansen en bedreigingen liggen voor Publitas.

Enquête customer succes managers

Tijdens het onderzoek is een online enquête verspreid onder customer success managers uit verschillende regio's om inzicht te krijgen in de verschillen in het gebruik van data. Dit is online uitgevoerd omdat het bedrijf volledig op afstand werkt en meerdere landen betreft. De enquête is ingevuld door zes customer success managers. Deze enquête bleek achteraf geen toegevoegde waarde te zijn voor deelvraag 1, maar bood wel een waardevol inzicht in de huidige stand van zaken rondom het gebruik van folderdata.

Literatuur

Er is literatuuronderzoek uitgevoerd ter onderbouwing van de vraag of de digitale folder wel de juiste keuze is voor de retailer en of deze markt in de toekomst zal groeien. De onderdelen die in de literatuur terugkomen zijn het GfK folderonderzoek, dit is een nationaal onderzoek naar reclamefolders, de effecten van corona, oorlog en grondstofprijzen.

3.2.1.2 Onderzoeksverloop deelvraag 1

In deze deelvraag is gestart met het uitvoeren van de interviews om in kaart te brengen wat de huidige strategie is en een goede basis te zetten voor het vervolg. Vervolgens zijn het 5-krachtenmodel van Porter en de SWOT-analyse ingevuld. Het 5-krachtenmodel is gevuld met onder andere eigen observaties en input vanuit de expertinterviews. Met eigen observaties bedoel ik korte gesprekken met werknemers en het zelf bezoeken van websites van concurrenten. Daarnaast is ook gebruikgemaakt van interviewdata. Vervolgens is met alle bovengenoemde data de SWOT-analyse opgesteld.

Het literatuuronderzoek en de enquêtes zijn opgesteld om meer inzicht te krijgen in de randzaken omtrent data en externe factoren die effect hebben op de foldermarkt. De enquêtes zijn door zes personen ingevuld, met enige moeite en navragen.

3.2.1.3 Betrouwbaarheid en validiteit deelvraag 1

Tijdens de uitvoering van deelvraag 1 werden betrouwbaarheid en validiteit gewaarborgd in de semigestructureerde interviews en enquêtes door dezelfde vragen op dezelfde dag aan de betrokken partijen te stellen. Er is ook gecontroleerd of de verkregen resultaten overeenkomen met de eigen observaties.

Na de dataverzameling vond een toetsingsmoment plaats bij de technisch directeur, waarbij de verzamelde data werd gepresenteerd, discussies plaats vonden en feedback werd gegeven. Deze feedback is vervolgens verwerkt in het onderzoek.

3.2.2 Deelvraag 2

Deelvraag 2 luidt als volgt: "Hoe ziet de huidige data-architectuur eruit bij zowel WePublish als Publitas en wat zijn de sterke punten en zwakke punten ervan?".

3.2.2.1 Dataverzameling

IT-Architectuurmodellering

Om antwoord te geven op de deelvraag is de IT-architectuur van WePublish en Publitas gemodelleerd door middel van de modelleringstaal Archimate. Dit is gedaan op basis van onder andere eigen observaties van de IT-architecturen, schematische tekeningen van de IT-architecturen en semi-gestructureerde interviews met een developer van WePublish en een werknemer uit het DevOps team van Publitas.

Semigestructureerde interviews

Tijdens het uitvoeren van deelvraag 2 zijn twee semigestructureerde interviews opgesteld en uitgevoerd. Deze interviews zijn gehouden met een werknemer uit het DevOps team van Publitas en een Developer bij WePublish. Er is voor het DevOps team gekozen omdat zij verantwoordelijk zijn voor de IT-architectuur van Publitas en voor een developer bij WePublish omdat er het afgelopen jaar geen data-engineer of DevOps aanwezig was bij WePublish. Development was in deze tijd verantwoordelijk voor de IT-architectuur. De Archimate-tekeningen zijn ook doorgenomen tijdens deze interviews.

De twee interviews hadden twee doelstellingen. Eén daarvan is het vaststellen van de sterke en zwakke punten van het systeem, waarbij wordt geëvalueerd of de datakwaliteit, schaalbaarheid en veiligheid zijn gewaarborgd. Het tweede doel is het valideren van de Archimate-tekeningen.

Literatuur

In deelvraag 2 is voornamelijk literatuur gebruikt om de vaktermen en specifieke onderdelen uit te leggen, te verduidelijken en/of te begrijpen.

3.2.2.2 Onderzoeksverloop deelvraag 2

Het verloop van deelvraag is als volgt: Als start zijn er schematische tekeningen opgevraagd van de IT-architecturen van beide organisaties. Ook zijn de IT-architecturen door middel van eigen observaties bekeken. Door deze informatie samen te voegen, is een eerste gemodelleerde IT-architectuurtekening gemaakt binnen Archimate. Vervolgens zijn er twee semigestructureerde interviews afgenomen met twee doeleinden. Eén daarvan is het in kaart brengen van de datakwaliteit, schaalbaarheid en veiligheid van de huidige IT-architectuur. Het tweede doel is het valideren van de Archimate-tekeningen.

3.2.2.3 Betrouwbaarheid en validiteit deelvraag 2

Om de betrouwbaarheid en validiteit te waarborgen, is data-triangulatie toegepast. Het hoofddoel van de deelvraag was het opstellen van een correct gemodelleerde IT-architectuur voor beide organisaties. Door verschillende onderzoeksmethodes hierop toe te passen, namelijk op basis van eigen observaties en schematische tekeningen, is een eerste model binnen Archimate gemaakt. Dit model is vervolgens gevalideerd door middel van semi-gestructureerde interviews, waardoor de betrouwbaarheid en validiteit gewaarborgd zijn.

3.2.3 Deelvraag 3

Deelvraag 3 luidt als volgt: "Hoe dient de geconsolideerde data van WePublish en Publitas vormgegeven te worden?".

3.2.3.1 Dataverzameling

Data-mapping/analyse

Om de huidige datastructuur van beide organisaties in kaart te brengen, is de methode van data-mapping toegepast. Hierbij wordt voornamelijk vanuit eigen observaties gewerkt met behulp van de Google Analytics 4 Debugger en de Snowplow Debugger. Hierdoor kan men door het uitvoeren van een handeling precies achterhalen welke data wordt meegestuurd. Vervolgens is er in de database ook gekeken naar de relevante tabellen en dit is allemaal genoteerd.

Stakeholder analyse

Om tot een toekomstig advies te komen, moeten de wensen van de stakeholders duidelijk zijn. Om vast te stellen wie de stakeholders zijn is een stakeholderanalyse uitgevoerd om ze te identificeren, categoriseren en prioriteren.

Semigestructureerde interviews

Tijdens het uitvoeren van deelvraag 3 zijn twee semigestructureerde interviews opgesteld en uitgevoerd. Deze interviews zijn gehouden met een Senior Retail Consultant en een Accountmanager/Sales. Er is voor hen gekozen omdat zij de primaire stakeholders zijn van het project. Tijdens deze interviews zijn de onderwerpen folder data tracking en dashboarding behandeld.

GAP-analyse

Om de verschillen tussen de huidige situatie en de gewenste situatie te onderzoeken, is een GAP-analyse opgesteld (Vleeshouwers, 2023). Deze analyse maakt gebruik van de resultaten van de eerdergenoemde methodes.

3.2.3.2 Onderzoeksverloop deelvraag 3

Het verloop van de deelvraag is als volgt: Door middel van de twee debuggers is de data in kaart gebracht die werd gemeten. Ook is er gekeken naar de bijbehorende tabellen in de MsSQL- en Databricks-databases. Nadat dit beschreven was, werd er een stakeholderanalyse uitgevoerd. Vervolgens werden op basis van deze uitkomsten twee semi-gestructureerde interviews opgesteld voor de twee primaire stakeholders. De resultaten van de bovengenoemde methoden worden vervolgens gebruikt voor de GAP-analyse.

3.2.3.3 Betrouwbaarheid en validiteit deelvraag 3

Tijdens de uitvoering van deelvraag 3 werd betrouwbaarheid en validiteit gewaarborgd in de semi-gestructureerde interviews door dezelfde vragen op dezelfde dag aan de betrokken partijen te stellen. Ook is er rekening gehouden met het feit dat alleen de primaire stakeholders werden geïnterviewd en niet de secundaire of minder belangrijke stakeholders.

3.2.4 Deelvraag 4

Deelvraag 4 luidt als volgt: "Hoe dient de toekomstige IT-architectuur eruit te zien voor Publitas, waarin zij de interne en externe folderdata kunnen opslaan?".

3.2.4.1 Dataverzameling

IT-architectuurmodellering

Om antwoord te geven op de deelvraag is de toekomstige IT-architectuur van WePublish en Publitas gemodelleerd door middel van de modelleringstaal Archimate. Dit is gedaan op basis van onder andere de resultaten uit eerdere deelvragen, de brainstorm en workshop meeting met de data-engineer van Publitas en literatuur.

Brainstorm sessie/Workshop sessie

Om het initiële ontwerp van de toekomstige IT-architectuur aan te scherpen en te valideren, is er een brainstorm/workshopsessie opgezet met de data-engineer. Dit is gedaan met de data-engineer omdat hij verantwoordelijk is voor de toekomstige IT-architectuur van Publitas en WePublish.

(Succes begint met een Brainstormsessie - Samen 100+ ideeën uitdenken?, 2023)

Literatuur

Om het ontwerp te valideren en alternatieve opties tegen elkaar af te wegen, is literatuur geraadpleegd. Hiervoor zijn zowel interne literatuur, zoals documentatie, als externe literatuur gebruikt.

3.2.4.2 *Onderzoeksverloop deelvraag 4*

Om tot de geconsolideerde toekomstige IT-architectuur van Publitas en WePublish te komen, is gebruik gemaakt van interne beschikbare kennis, kennis dat is opgedaan tijdens deelvragen 2 en 3 en literatuur over IT-architectuur en IT-tooling. Om tot een eerste ontwerp te komen, is er een ontwerp gemaakt op basis van het bovenstaande binnen Archimate. Om dit te valideren en het ontwerp aan te scherpen, is er een interactieve workshop/brainstorm sessie gehouden met de data-engineer van Publitas. Tijdens deze sessie werd het huidige ontwerp getoond en werd informatie ingewonnen over zijn specialisatie in tooling, ervaring en algemene kennis. Met alle bovenstaande informatie is het initiële ontwerp aangescherpt en is het technisch ontwerp uitgewerkt en is deelvraag 4 beantwoord.

3.2.4.3 *Betrouwbaarheid en validiteit deelvraag 4*

Door eerst een initieel ontwerp op te stellen op basis van de voorgaande deelresultaten, de documentatie van de organisatie en externe literatuur en daarna pas met de data-engineer te spreken, is de betrouwbaarheid gewaarborgd. Dit komt doordat het gesprek niet voorafgaand aan de initiële tekening heeft plaatsgevonden, waardoor zijn senioriteit en kennis dit niet heeft beïnvloed. De tekeningen zijn vervolgens wel aangescherpt op basis van de input van de data-engineer. Hiervoor zijn wel afwegingen gemaakt, waarbij is gekeken naar wat goed is voor de organisatie.

3.2.5 Deelvraag 5

Deelvraag 5 luidt als volgt: "Hoe kan de geconsolideerde data effectief worden gevisualiseerd om waardevolle inzichten te bieden aan de stakeholders?".

3.2.5.1 *Dataverzameling*

Stakeholderanalyse

Tijdens de uitvoering van deelvraag 5 is gebruikgemaakt van de stakeholderanalyse die eerder is uitgevoerd in deelvraag 3.

Semi-gestructureerde interviews

Tijdens de dataverzameling van deelvraag 5 is geconcludeerd dat er extra vragen gesteld moeten worden aan de primaire stakeholders. Hiervoor zijn nieuwe vragen opgesteld die specifieker ingaan op het dashboard en het ontwerp ervan en is een nieuw interview gepland en gehouden met de primaire stakeholders. De interviewtranscripties zijn vervolgens aangevuld met de nieuwe informatie. De primaire stakeholders zijn de senior retail consultant en de sales/accountmanager.

Requirementsanalyse

Op basis van het voorgaande is een requirementsanalyse opgesteld. In deze analyse wordt beschreven wat wel en niet in het dashboard moet worden opgenomen en wat de randvoorwaarden zijn om tot een succesvol dashboard te komen. Hierbij is ook een prioritering opgenomen volgens de MoSCoW-methode.

Wireframes en literatuuronderzoek

Op basis van de requirement analyse zijn wireframes opgesteld die het toekomstige ontwerp laten zien. Naast het rekening houden met de requirementsanalyse, is er ook aandacht besteed aan de literatuur van Stephan Few. De literatuur van Stephan Few in het theoretisch kader beschreven.

3.2.5.2 *Onderzoeksverloop deelvraag 5*

"Om tot een eerste ontwerp te komen voor de interne en externe gegevens, zijn de belanghebbenden die naar voren kwamen in deelvraag 3 opnieuw geïnterviewd om extra informatie te verzamelen over het dashboard.

Met zowel de oude als de nieuw verkregen informatie is vervolgens een requirementsanalyse opgesteld en geprioriteerd via de MoSCoW-methode.

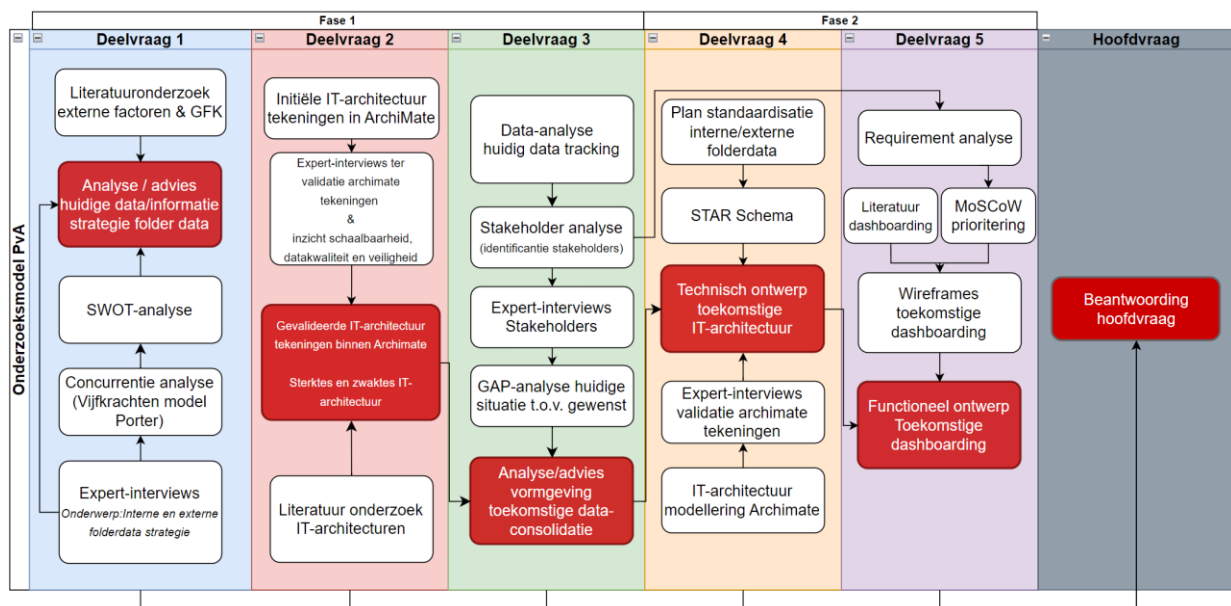
Op basis van al het bovenstaande zijn vervolgens verschillende wireframes voor het dashboard opgezet, rekening houdend met de dashboardliteratuur van Stephan Few. Ook zijn de wireframes op hoog over niveau beschreven, inclusief specifieke elementen.

3.2.5.3 *Betrouwbaarheid en validiteit deelvraag 5*

Om de betrouwbaarheid en validiteit te waarborgen en zo een goed dashboard op te leveren, is vanuit verschillende invalshoeken naar het dashboard gekeken. De eerste invalshoek is de requirementsanalyse, die intern is gemotiveerd vanuit interviews. Om dit vervolgens te toetsen, is de literatuur van Stephan Few geraadpleegd met betrekking tot dashboards. Tot slot is het dashboard getoetst bij de technisch directeur van WePublish, de eindverantwoordelijke van het project en een accountmanager om te bevestigen of de wireframes aan de behoefte voldoen.

3.3 Onderzoeksmodel

In dit hoofdstuk wordt het onderzoeksmodel getoond en wordt de rode draad van het onderzoek toegelicht. Zie hieronder het onderzoeksmodel:



Figuur 5 Onderzoeksmodel

Rode draad van het onderzoek

Hieronder wordt beknopt de rode draad van het onderzoek toegelicht. Het onderzoek kan in twee fasen worden opgedeeld: de eerste fase betreft het onderzoeken van de huidige situatie en de tweede fase omvat de toekomstige situatie. Hieronder volgt verdere toelichting.

Fase 1

Om de hoofdvraag te beantwoorden en Publitas een ontwerp/advies te bieden over hoe zij in de toekomst de twee IT-architecturen/datatoepassingen effectief kunnen consolideren, is eerst de huidige staat van beide partijen geanalyseerd. Dit is op drie niveaus gedaan. In deelvraag 1 is op strategisch niveau gekeken naar de datavoorziening en naar de toekomst. Kortom, wat doen beide organisaties nu en wat is het advies voor de toekomst? Vervolgens is in deelvraag 2 de IT-architectuur van beide partijen in kaart gebracht en gemodelleerd binnen Archimate en tot slot in deelvraag 3 het derde niveau, de data. Dus welke data wordt getrackt, welke verschillen zijn er tussen de organisaties en hoe dient dit in een toekomstige situatie eruit te zien. Bovenstaande kan worden gezien als fase 1 van het onderzoek.

Fase 2

Fase 2 begint bij deelvraag 4. Met inachtneming van de kennis die is opgedaan uit deelvragen 1 t/m 3, samen met aanvullend onderzoek en expertinterviews, is een ontwerp opgesteld voor een geconsolideerde IT-architectuur, evenals advies voor de interne en externe data. Deze bevindingen worden opgenomen in het technisch ontwerp. Vervolgens zijn in deelvraag 5 de wireframes opgesteld die aantonen hoe een toekomstig dashboard eruit kan zien, gebaseerd op de input van de primaire stakeholders en literatuur van Stephan few. Dit wordt opgenomen in een functioneel ontwerp. Al het bovenstaande wordt als input gebruikt om antwoord te geven op de hoofdvraag van het onderzoek.

4 Resultaten

In dit hoofdstuk worden alle onderzoeksresultaten per deelvraag besproken, verder wordt voor elke deelvraag een aparte deelconclusie uitgewerkt. De beantwoording van de hoofdvraag vindt plaats in het hoofdstuk Conclusie.

4.1 Deelvraag 1

In dit hoofdstuk worden de resultaten besproken van de volgende deelvraag: “Wat zijn de consequenties van de fusie voor de informatievoorziening strategie met betrekking tot folderdata van Publitas en welk advies is er voor de toekomstige strategie omtrent de informatievoorziening?”.

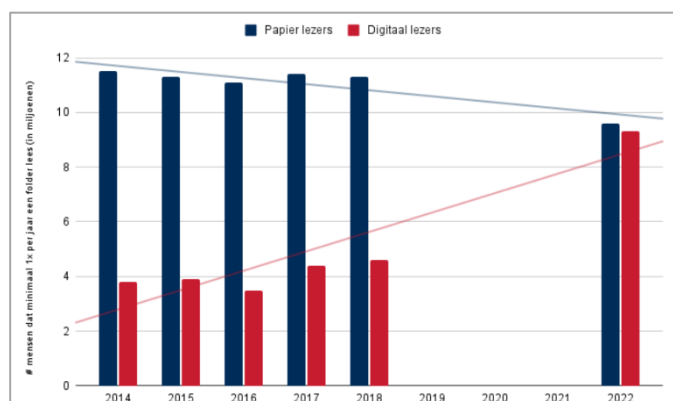
4.1.1 Resultaten

Uit het interview met de technisch directeur van WePublish is gebleken dat WePublish de folderdata momenteel voor drie doeleinden gebruikt in de huidige strategie. Dit wordt ingezet voor interne monitoring en schaalbaarheid, om inzicht te krijgen in de prestaties van de digitale folder en voor het vergelijken van de prestaties tussen de verschillende platformen waar de digitale folder op staat. Uit het interview blijkt dat dit punt tevens het belangrijkste is.

Verder bleek uit het interview dat de CEO van Publitas aanvankelijk overwoog het folderplatform Folders.nl te verkopen, maar nu de meerwaarde ervan inziet en er dus verder in investeert. Kortom, de focus wordt breder gelegd op het aanbod van folders. Uit het interview blijkt dat WePublish zich gaat richten op affiliate platformen, terwijl Publitas zich zal concentreren op de eigen retailer platformen. Daarom ziet de technisch directeur ook geen grote consequenties wat betreft de strategie en visie en merkt een duidelijke scheiding op: Publitas behandelt eigen platformen, terwijl WePublish zich richt op de affiliate platformen (volledige interviewtranscript is opgenomen in Bijlage 1).

Uit het interview met de Data en CRM-lead van Publitas blijkt dat Publitas een andere benadering heeft als het gaat om de strategie rondom folderdata. Publitas slaat de data op om één waarheid te hebben. Wanneer er wordt gekeken naar externe folderdata, ondersteunt Publitas de retailer door inzichten te bieden in de vorm van KPI's, die vervolgens worden gevisualiseerd in een dashboard. De data en CRM-lead geeft echter aan dat dit een nieuw aspect is. Publitas is van plan dit verder te ontwikkelen en ziet ook de positieve aspecten van WePublish haar visie met betrekking tot de folderplatformen. Niettemin erkennen ze dat ze de expertise van WePublish nodig hebben om het Publitas-dashboard verder te ontwikkelen (volledige interviewtranscript is opgenomen in Bijlage 2).

Uit het folderonderzoek van het marktonderzoeksinstituut GfK blijkt dat de digitale folder een stevige opmars maakt ten opzichte van de papieren folder. De papieren folder wordt nog steeds het meest gelezen met 9.6 miljoen regelmatige gebruikers, terwijl de opkomende digitale folder met 9.3 miljoen lezers goed scoort. Dit is een verdubbeling ten opzichte van 2018 (GfK, 2022).



Figuur 6 Grafiek GfK

Wanneer er gekeken wordt naar de vier facetten van de SWOT-analyse wordt duidelijk dat de fusie aanzienlijke kennis heeft samengebracht met betrekking tot de digitale folder. Op basis van de SWOT-analyse kan geconcludeerd worden dat het van cruciaal belang is voor Publitas en WePublish om in eerste instantie de processen, tools en architectuur te consolideren. Dit legt de basis voor het benutten van de volgende grote kansen.



Figuur 7 SWOT-analyse

De toekomstige kansen liggen

voornamelijk op het gebied van data.

Voor Publitas is het cruciaal om niet

uitsluitend afhankelijk te zijn van het publiceren van digitale folders. Het is essentieel dat retailers, naast het publiceren van deze folders, ook de aanvullende waarde van Publitas erkennen. De aanvullende waarde zal volgens de technisch directeur het faciliteren van de verspreiding van digitale folders naar diverse folderplatformen moeten zijn en het verstrekken van diepgaande data-inzichten op zowel het eigen platform als op de folderplatformen en het ondersteunen van retailers bij de inhoud van de digitale folder met behulp van machine learning. De CEO van Publitas deelt deze mening met de technisch directeur van WePublish. De volledige SWOT-Analyse is terug te vinden in bijlage 4.

4.1.2 Conclusie deelvraag 1

“Wat zijn de consequenties van de fusie voor de informatievoorziening strategie met betrekking tot folderdata van Publitas en welk advies is er voor de toekomstige strategie omtrent de informatievoorziening?”

De fusie tussen WePublish en Publitas heeft aanzienlijke gevolgen voor de informatievoorziening strategie met betrekking tot folderdata. Publitas heeft nu zijn informatiestrategie rondom folderdata gericht op het bieden van één waarheid en basisinzichten voor de retailer via dashboards. Het tweede aspect betreft de expertise van WePublish, waar zij veel ervaring in meebrengen. Bovendien blijkt uit het GFK-onderzoek dat er nog voldoende groeimogelijkheden zijn in de digitale foldermarkt en dat hierin nog geïnvesteerd kan worden. Verder komt duidelijk naar voren dat Publitas zich richt op het eigen platform van de retailer, terwijl WePublish zich richt op een combinatie van het eigen platform en affiliate folderplatformen. De visie van WePublish is echter sterk gericht op de affiliate folderplatformen, wat ook door de CEO van Publitas, als een positieve ontwikkeling wordt beschouwd.

In de bovenstaande alinea is de informatiestrategie met betrekking tot folderdata beschreven.

Echter, zowel Publitas als WePublish hebben de strategie niet zelf uitgeschreven met een pakkende zin of korte alinea. Dit komt doordat het slechts een onderdeel is van de bredere organisatiestrategie die buiten de scope van dit onderzoek valt.

Het advies rondom de strategie is tweeledig. Publitas moet gebruikmaken van de expertise op het gebied van data-visualisatie en folderoptimalisatie vanuit WePublish en zich richten op de digitale folder op de eigen platformen van de retailer en de bijbehorende dashboards. Tegelijkertijd zou de WePublish-tak zich moeten richten op de affiliate folderplatformen, het bouwen van koppelingen tussen deze platformen en de data-visualisatie en dashboards ervan.

4.2 Deelvraag 2

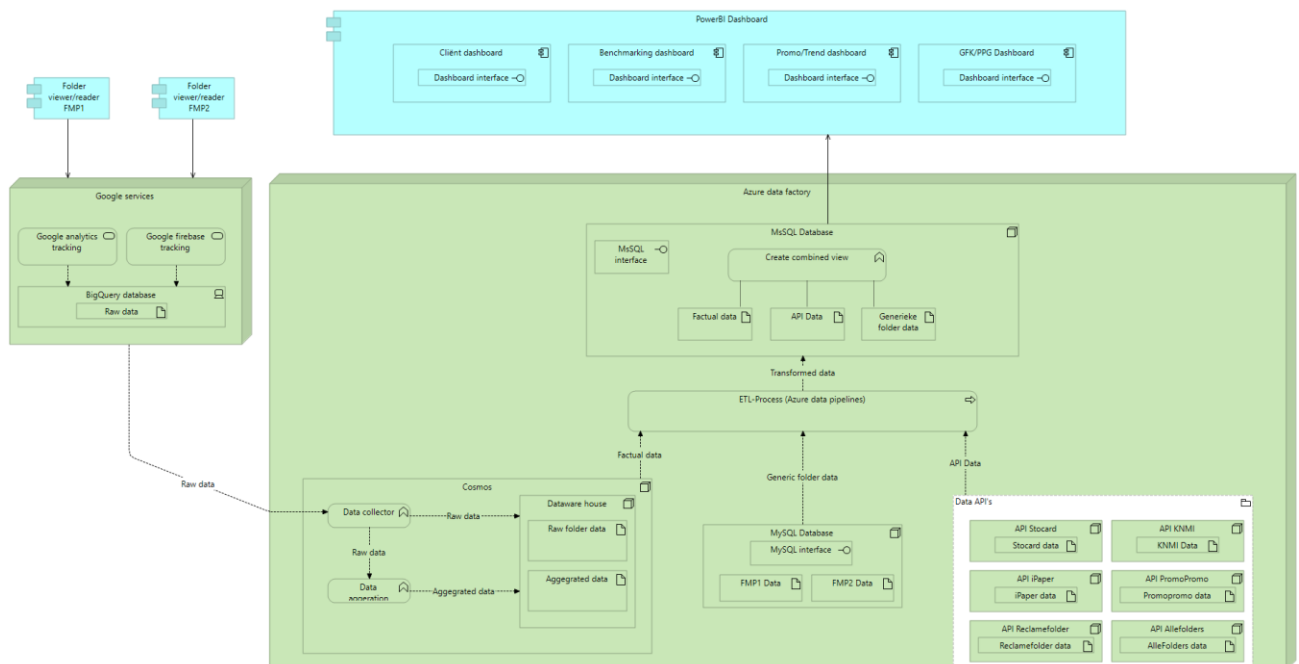
In dit hoofdstuk worden de resultaten besproken van de volgende deelvraag: “Hoe ziet de huidige data-architectuur eruit bij zowel WePublish als Publitas en wat zijn de sterke punten en zwakke punten ervan?”.

4.2.1 Resultaten

In dit hoofdstuk worden de IT-architecturen apart toegelicht.

4.2.1.1 WePublish IT-architectuur

Zie hieronder de gemodelleerde IT-architectuur van WePublish. Onder de afbeelding is een beknopte uitleg opgenomen van de situatie en een gedetailleerde beschrijving van het onderstaande is te vinden in bijlage 7.



Figuur 8 WePublish IT-architectuur

De WePublish IT-architectuur is inzichtelijk gemaakt aan de hand van Archimate en omvat drie facetten. Deze facetten zijn als volgt gedefinieerd:

Data Tracking en Opslag: Deze activiteiten vinden plaats in de twee folder viewers en worden getrackt via Google-services. De getrackte data wordt opgeslagen in BigQuery.

Azure Data Factory: Binnen deze omgeving worden verschillende taken uitgevoerd. Cosmos haalt ruwe data op vanuit BigQuery en slaat deze op. Daarnaast bevindt zich de MySQL-database, waarin generieke data van het foldermanagementplatform wordt opgeslagen, samen met de Data API's. Al het bovenstaande wordt opgehaald via een ETL-proces en opgeslagen in de MySQL-database.

PowerBI Dashboard: Tot slot wordt de opgeslagen data opgehaald door het PowerBI-dashboard, waar het wordt gevisualiseerd.

Uit het interview met Toby Nijboer blijkt dat de IT-architectuur van WePublish slecht wordt onderhouden en dat de datakwaliteit niet wordt gewaarborgd. Echter, pas wanneer een data-analist foutieve cijfers tegenkomt, wordt er actie ondernomen. Verder is er minimaal nagedacht over schaalbaarheid bij de ontwikkeling. De feiten tabel met de gedragsdata blijft momenteel groeien en

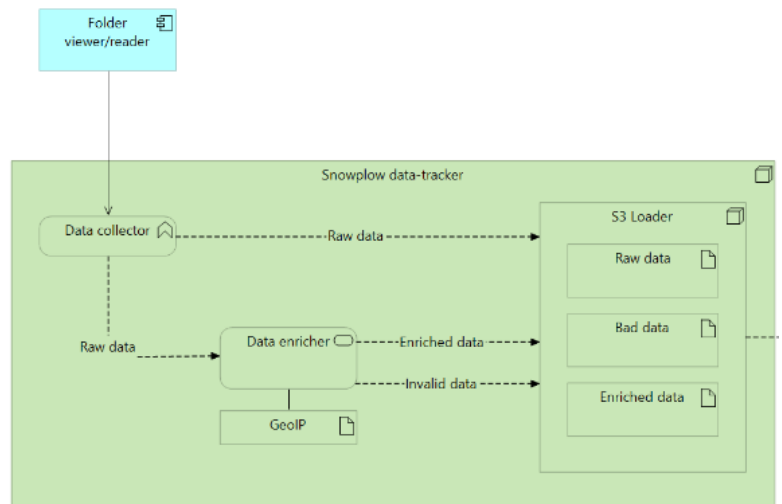
moet handmatig aangepast worden als er problemen ontstaan, voornamelijk doordat er met statische waarden zoals datumranges wordt gewerkt. Qua databeveiliging heeft WePublish het wel op orde. Databases zijn goed afgeschermd aan de hand van IP-whitelisting, gebruikersnamen en wachtwoorden en er is 2FA ingesteld voor de Azure-omgeving. Verder heeft Toby Nijboer de Archimate tekening gecontroleerd (volledige interviewtranscript is opgenomen in Bijlage 5).

4.2.1.2 Publitas IT-architectuur

De Publitas IT-architectuur, die gehost wordt op Amazon Web Services, is inzichtelijk gemaakt aan de hand van Archimate en omvat twee facetten. Deze facetten zijn als volgt gedefinieerd:

Data tracking

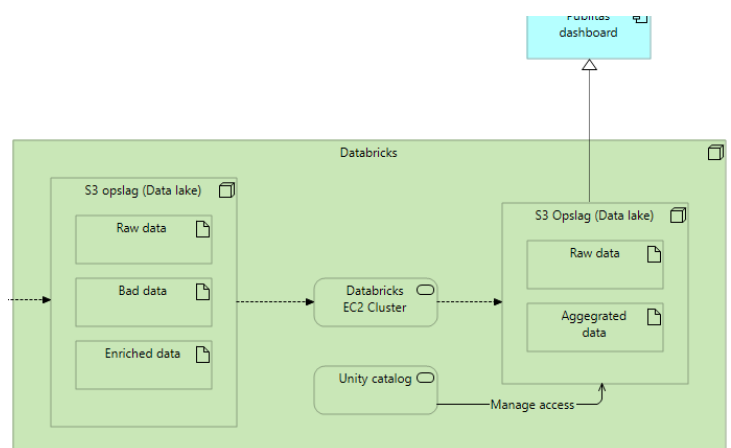
Binnen de applicatie laag van Publitas is de folder viewer van Publitas opgenomen. Van deze viewer worden de gegevens getrackt met de Snowplow-data tracker. De gegevens komen binnen in Snowplow via de data collector en de onbewerkte ruwe data stroomt door naar de S3 Loader. Daarnaast loopt een lijn naar de data enricher, waar de gegevens worden verrijkt met GeoIP en gecheckt aan de hand van een schema registry. Vervolgens loopt deze lijn ook door naar de S3 Loader. In de S3 Loader worden de ruwe data, bad data en verrijkte data verwerkt en doorgestuurd naar een S3 opslag binnen Databricks.



Figuur 10 Publitas data tracking

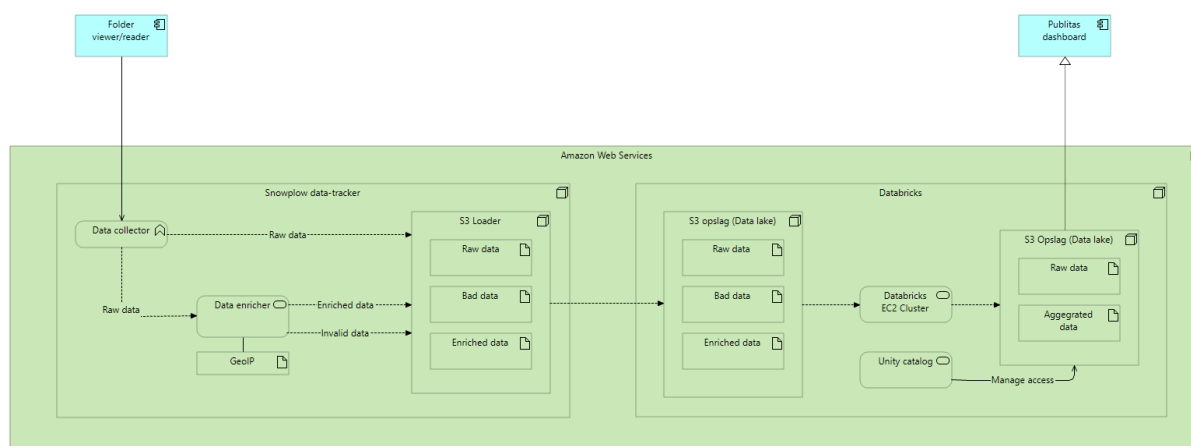
Dataopslag

De data komt binnen in de S3-opslag in Databricks vanuit de S3 Loader. Deze data wordt vervolgens verwerkt en geaggregeerd via een ETL-proces via de EC2 Clusters en opnieuw opgeslagen in een S3-opslag (Data Lake). De Unity-catalogus heeft vervolgens een koppeling met de S3 opslag. Daarnaast loopt de lijn ook naar het Publitas-dashboard, dat ook gebruik maakt van deze gegevens. Publitas heeft één extern dashboard.



Figuur 9 Publitas dataopslag

Volledig gemodelleerde architectuur



Figuur 11 Publitas IT-architectuur

Uit het interview met Umer Khan blijkt dat Publitas de kwaliteit van de getrackte data waarborgt binnen Snowplow. Publitas slaat de ruwe data op maar heeft ook een datalijn die wordt verrijkt. Deze verrijkte datalijn loopt ook langs het proces van een schema registry. Deze registry valideert de data door te controleren of deze de juiste waardes bevat.

Publitas waarborgt ook de schaalbaarheid door allereerst het systeem zo in te richten dat het 3 tot 4 keer de huidige workload aankan. Bovendien worden alle facetten gemonitord op load capacity. Wanneer de drempel wordt overschreden, ontvangt het DevOps-team automatisch een melding via Slack en schalen ze het systeem op indien nodig.

Voor de databeveiliging is een Virtual Private Cloud (VPC) opgezet binnen Amazon Web Services. In de bovenstaande modulatie valt de VPC dus onder de Amazon Web Services-node. De VPC zorgt ervoor dat de services en componenten niet van buitenaf bereikbaar zijn en dit alleen te bereiken is via de VPN van Publitas. Publitas heeft ook een concept geïmplementeerd dat ze 'reducing the blast radius' noemen. Dit houdt in dat wanneer één punt geïnfilteerd is, dit geen effect heeft op de rest. Bovendien is Publitas alert op updates en maakt het gebruik van IP-whitelisting en GitHub secrets voor wachtwoordmanagement (volledige interviewtranscript is opgenomen in Bijlage 6).

4.2.2 Conclusie deelvraag 2

Hoe ziet de huidige data-architectuur eruit bij zowel WePublish als Publitas en wat zijn de sterke punten en zwakke punten ervan?

Hoe de IT-architectuur van beide partijen eruit ziet, is beknopt weergegeven in bovenstaand hoofdstuk. In bijlage 7 vind je een uitgebreidere beschrijving van de WePublish-architectuur.

Bij het beoordelen van de sterke en zwakke punten van de IT-architecturen kan worden geconcludeerd dat de IT-architectuur van WePublish tekortschiet. De datakwaliteit wordt niet gewaarborgd en ook is de schaalbaarheid niet goed ingeregeld, dit moet namelijk handmatig worden aangepast. Verder ontbreekt het aan monitoring om dit vroegtijdig op te vangen. Desalniettemin is de beveiliging wel goed ingeregeld.

Wanneer vervolgens naar de IT-architectuur van Publitas wordt gekeken, kan worden geconcludeerd dat de datakwaliteit goed is, omdat deze wordt gewaarborgd door middel van een validatieproces binnen een schema register. Ook is de schaalbaarheid goed ingeregeld, omdat het systeem 3 tot 4 keer de huidige workload aankan, wat ook wordt gemonitord met automatische meldingen. Tot slot is de databeveiliging ook op orde, onder andere door whitelisting en 2FA.

4.3 Deelvraag 3

In dit hoofdstuk worden de resultaten besproken van de volgende deelvraag: “Hoe dient de geconsolideerde data van WePublish en Publitas vormgegeven te worden?”

4.3.1 Resultaten

Uit de resultaten van de data-mapping blijkt dat over het algemeen dezelfde data wordt bijgehouden vanuit de digitale folder door zowel Publitas als WePublish. Er zijn echter gedetailleerde verschillen waar te nemen. Wanneer er in de Publitas viewer een interactie wordt aangeklikt en er verschijnt een nieuw venster of een pop-up binnen de folder en deze vervolgens wordt gesloten, wordt er een nieuwe pageview getrackt. WePublish doet dit niet, zij meten slechts één pageview wanneer een pagina wordt bezocht.

Verder, wanneer een spread in de folder viewer wordt bekeken (spread zijn 2 pagina's naast elkaar), worden de events op deze pagina's opgeslagen onder bijvoorbeeld 'page_number: 2-3' en niet specifiek op de individuele pagina's waar dit plaatsvindt. Echter is het mogelijk om aan de hand van coördinaten en het ID van de interactie te achterhalen op welke pagina dit heeft plaatsgevonden. Na eigen observaties van Asana-taken en Slack-berichten blijkt dat dit punt al is aangekaart en er zijn plannen om dit in de toekomst op te splitsen, in bijlage 8 is de data-mapping analyse opgenomen van WePublish en Publitas.

Ibe Hundling werkt bij Publitas als senior retail consultant en is één van de primaire stakeholders, omdat hij actief met data en de retail samenwerkt om de folder naar het volgende niveau te tillen (stakeholder analyse is in bijlage 9 opgenomen). Ibe heeft tijdens het interview aangegeven tevreden te zijn met de functionaliteiten van het WePublish-dashboard, maar minder tevreden te zijn over de continuïteit. Hiermee doelt Ibe op de bugs en vele onduidelijkheden voor de retailers. Daarnaast beschouwt hij het Publitas-dashboard als te basaal. Ibe vindt de manier waarop WePublish omgaat met pageviews goed. Verder geeft hij aan dat de opmaak verouderd is en er gekeken moet worden naar Publitas huisstijl. Ibe Hundling geeft de voorkeur aan Staaf- en lijndiagrammen en buttons als goede visualisaties, terwijl hij minder enthousiast is over cirkeldiagrammen (volledige interviewtranscript is opgenomen in Bijlage 10).

Ramon Eerdman werkt bij Publitas als sales medewerker en is één van de primaire stakeholders omdat hij actief data inzet om retailers binnen te halen en te upsellen (stakeholder analyse is in bijlage 9 opgenomen). Hij heeft tijdens het interview aangegeven tevreden te zijn met de functionaliteiten van de WePublish dashboarding, maar hij voert geen diepgaande analyses uit en betreft hiervoor het data team. Hij deelt verder de mening van de senior retail consultant dat de huisstijl verouderd is omdat deze nog steeds in de WePublish-huisstijl staat en niet in de Publitas-huisstijl. Wat voor Ramon echter van groot belang is, is dat WePublish zich verbreedt in het aanbod door externe folderplatforms te koppelen. Hij gelooft dat dit de waarde van WePublish voor de retailsector kan verhogen. Wat betreft de opmaak hecht Ramon waarde aan duidelijke visualisaties, zonder specifieke voorkeur over hoe dit wordt vormgegeven (volledige interviewtranscript is opgenomen in Bijlage 11).

Met de verzamelde data tot nu toe is een GAP-analyse uitgevoerd, wat heeft geleid tot vier concrete punten voor de gewenste (SOLL) situatie. Ten eerste is het wenselijk dat de WePublish-klanten migreren naar de Publitas backend, waarbij Publitas het beheer van de digitale folder op Retail platforms overneemt. Hierdoor zou de data-tracking moeten verlopen via de Publitas-architectuur, met gebruikmaking van Snowplow. Publitas dient de data-tracking aan te passen naar de eerder

beschreven gewenste situatie. WePublish zal zich dan voornamelijk gaan richten op folderplatformen en de data-visualisatie hiervan.

Daarnaast is het wenselijk om een gezamenlijke strategie te ontwikkelen met betrekking tot consultancy en sales met betrekking tot data en data-dashboarding. De senior retail consultant geeft aan dat Publitas momenteel kansen laat liggen op dit gebied. Tot slot dient de huisstijl van WePublish aangepast te worden naar die van Publitas. De volledige GAP-Analyse is terug te vinden in bijlage 12.

4.3.2 Conclusie deelvraag 3

“Hoe dient de geconsolideerde data van WePublish en Publitas vormgegeven te worden?”

Publitas dient de data-tracking aan te passen conform de wensen van de primaire stakeholders. Deze wensen komen overeen met de manier waarop WePublish momenteel de data trackt. Dit houdt in dat pageviews alleen moeten worden getrackt wanneer er van pagina wordt gewisseld en pagina-interacties specifiek moeten worden getrackt per pagina, niet op spread-niveau.

Voor de toekomstige situatie is het mogelijk om de historische data van Publitas en WePublish te consolideren. Hiervoor moet echter een data-oplossing worden ontwikkeld waarbij aan de hand van de interactiecoördinaten wordt vastgesteld op welke pagina van een spread is geklikt. De overkoepelende KPI's zoals unique views en clicks kunnen worden gecombineerd. Het is echter essentieel om aan de hand van goede dashboard filters een selectie te maken, om te vermijden dat er ongelijksoortige zaken met elkaar worden vergeleken. Pageviews daarentegen dienen niet te worden vergeleken binnen een dashboard, omdat hier twee verschillende trackingmethoden achter zitten.

4.4 Deelvraag 4

Nu de informatiestrategie omtrent folderdata is onderzocht, de huidige IT-architectuur en data in kaart zijn gebracht, wordt er gekeken naar de toekomstige situatie.

In dit hoofdstuk worden de resultaten besproken van de volgende deelvraag: “Hoe dient de toekomstige IT-architectuur eruit te zien voor Publitas, waarin zij de interne en externe folderdata kunnen opslaan?”

4.4.1 Resultaten / Technisch ontwerp

Nu Publitas en WePublish samen doorgaan als één organisatie, dient de IT-architectuur geconsolideerd te worden tot één samenhangend geheel dat het behalen van de missie en visie in de toekomst mogelijk maakt. In dit technisch ontwerp wordt een toekomstige IT-architectuur ontworpen en beschreven die hieraan moet bijdragen. De scope van het TO gaat specifiek over de IT-architectuur achter de folderdata. Verder is vanuit het management al besloten dat de WePublish klanten zullen worden gemigreerd naar de Back-end van Publitas en dus naar Publitas play. De voornaamste reden hiertoe is dat het klantenbestand velen malen groter is van Publitas ten opzichte van WePublish.

Bedrijfscontext IT-architectuur

In de huidige IT-architectuur wordt de folderdata voor diverse doeleinden ingezet. De technisch directeur (WePublish) geeft aan dat folderdata wordt ingezet voor interne monitoring en schaalbaarheid, om inzicht te verkrijgen in de prestaties van de digitale folder en voor het vergelijken van prestaties tussen verschillende platformen. Aan de andere kant benadrukt de data en CRM-lead (Publitas) dat het één waarheid moet bieden en de retailer inzicht moet verschaffen in de prestaties. (volledige interviewtranscripten zijn opgenomen in Bijlage 1 en 2).

In hoofdstuk 4.1.2 is het volgende advies gegeven omtrent de toekomstige strategie “Publitas moet gebruikmaken van de expertise op het gebied van data-visualisatie en folderoptimalisatie vanuit WePublish en zich richten op de digitale folder op eigen platformen en de bijbehorende dashboards. Tegelijkertijd zou de WePublish-tak zich moeten richten op de affiliate platformen, het bouwen van koppelingen tussen deze platformen en de data-visualisatie en dashboards ervan.” Het is dus van belang dat de toekomstige IT-architectuur dit mogelijk maakt.

Kort samengevat: huidige IT-architectuur

De huidige IT-architectuur is tweedelig, zowel WePublish als Publitas hebben een eigen IT-architectuur. WePublish heeft een IT-architectuur die is opgebouwd binnen Microsoft Azure. Binnen deze architectuur worden verschillende gegevensbronnen samengebracht, waaronder de folderdata vanuit de viewer, folderdata van externe platformen en data van bijvoorbeeld het KNMI. Voor een complete uitleg van de huidige IT-architectuur van WePublish zie bijlage 7.

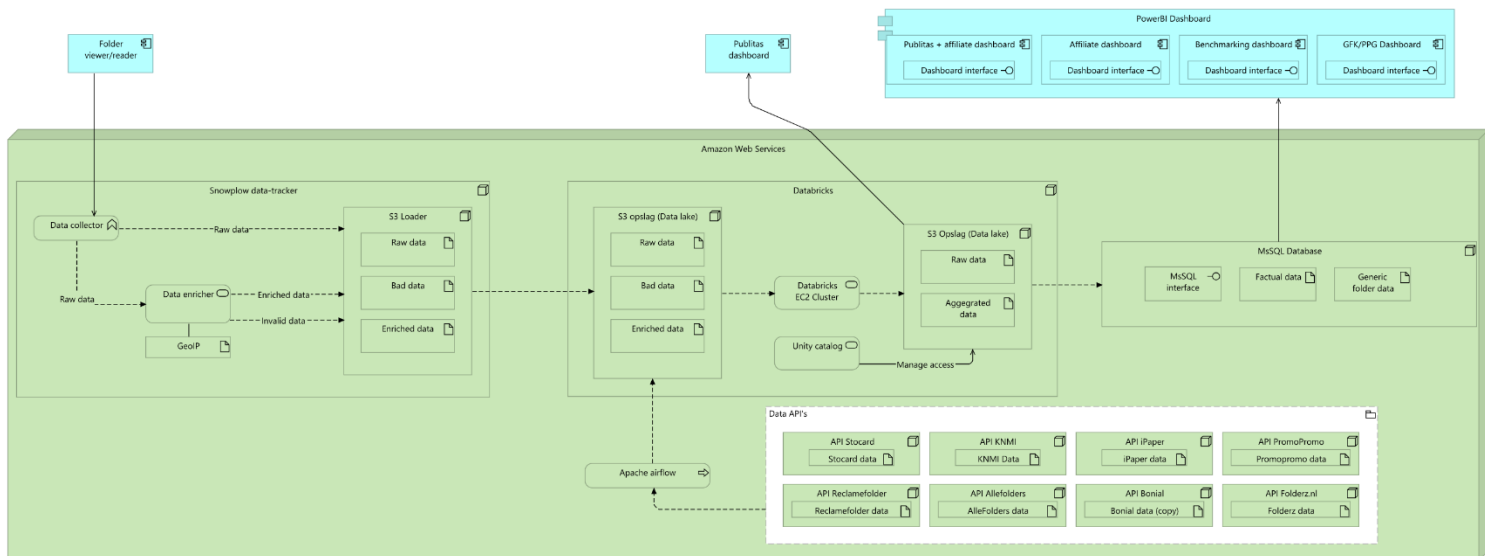
Publitas heeft daarnaast een IT-architectuur die is opgebouwd binnen AWS. Hier wordt de getrackte data van de folder viewer verwerkt en opgeslagen. Binnen Publitas Play wordt deze data gevisualiseerd. Publitas heeft geen verdere externe koppelingen met bijvoorbeeld folderplatformen of het KNMI. Voor een complete uitleg van de huidige IT-architectuur van Publitas, zie hoofdstuk 4.2.1

Wat betreft datakwaliteit, schaalbaarheid en beveiliging, zijn er verschillende punten waarop WePublish momenteel tekortschiet. Zo is de datakwaliteit niet goed gewaarborgd en hetzelfde geldt voor de schaalbaarheid. Deze aspecten zijn bij Publitas wel goed ingeregeld. Beide organisaties

hebben de beveiliging op orde. Voor een complete beschrijving van de datakwaliteit, schaalbaarheid en beveiliging, zie hoofdstuk 4.2.1

Gewenste IT-architectuur

Zie hieronder de modellering van de gewenste toekomstige IT-architectuur van Publitas/WePublish. Onder de afbeelding van de IT-architectuur is een volledige beschrijving te vinden van de verschillende facetten.



Figuur 12 Gewenste IT-architectuur

Amazon Web Services

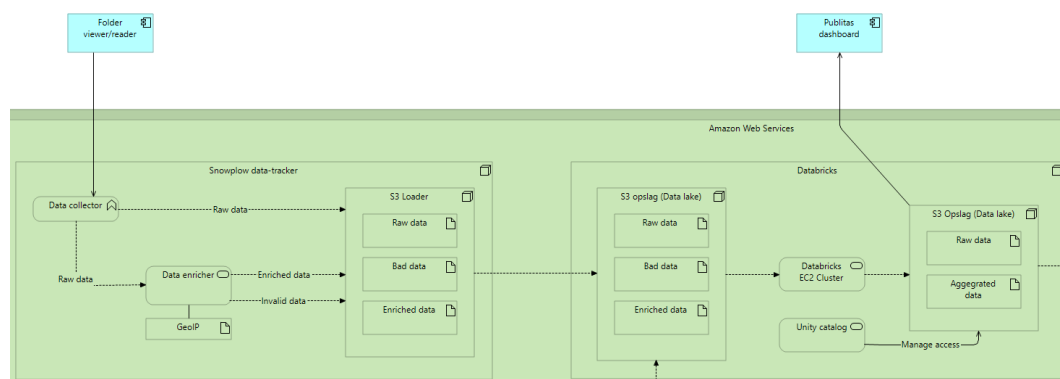
De toekomstige situatie zal worden gehost binnen AWS en niet in Azure. Hiervoor is gekozen om de volgende redenen:

- De gehele IT-architectuur van Publitas draait in AWS en de beslissing om de WePublish-klienten over te zetten naar de Publitas back-end is al genomen. Daarom is het onlogisch en niet efficiënt om vervolgens over te stappen naar Azure of een andere provider.
- Publitas beschikt over aanzienlijk meer data in AWS dan WePublish in Azure en de bijbehorende IT-governance is goed ingeregeld door Publitas, in tegenstelling tot WePublish.
- Het behouden van een hybride systeem is kostentechnisch niet efficiënt vanwege de grote hoeveelheid data die heen en weer moet worden gestuurd.

IT-architectuur Snowplow & Databricks:

De Snowplow-datatracking opzet verandert niet. Uit deelvraag 2 bleek dat deze opzet naar behoren functioneert en dat de datakwaliteit, schaalbaarheid en beveiliging goed zijn ingeregeld.

Wat wel verandert, is dat de WePublish-kanten gemigreerd worden naar de Publitas-back-end, genaamd Publitas Play. Deze keuze is intern al gemaakt. Wanneer er gekeken wordt naar de onderliggende architectuur en dan specifiek naar de onderwerpen datakwaliteit/tracking, schaalbaarheid en databeveiliging, blijkt dit de juiste keuze te zijn. Dit komt doordat het WePublish-systeem functioneert, maar de kwaliteit en schaalbaarheid niet goed zijn ingeregeld. De volledige uitleg over het bovenstaande is te vinden in hoofdstukken 4.2.1.1 en 4.2.1.2. Deze migratie houdt in dat de data van de WePublish-kanten in de toekomst wordt getrackt via Snowplow en wordt opgeslagen in Databricks.



Figuur 13 IT-architectuur Snowplow & Databricks

IT-architectuur API Data & MsSQL

In de toekomstige situatie wordt de data van externe kanalen opgehaald via API's. In de huidige situatie is dit niet altijd het geval, het komt bijvoorbeeld voor dat de data wordt aangeleverd door middel van Excel-sheets. De data wordt opgehaald door middel van Apache Airflow (What is Airflow™? — Airflow Documentation, sd) en dit stuurt de data door naar de S3-opslag van Databricks. Vervolgens wordt de API-data geaggregeerd en opgeslagen via de EC2-clusters in de S3-opslag.

Om de data op te halen is de afweging gemaakt om dit te doen doormiddel van Databricks of Apache Airflow. Hier is de keuze gemaakt voor Apache Airflow om de volgende redenen:

- Doordat Apache Airflow code-based is en niet over een gebruikersinterface beschikt, kan het als lastig worden ervaren om het te configureren. Echter heeft Publitas een Data Engineer, met veel ervaring met Apache Airflow.
- Apache is code based en open-source, waardoor er een reële kans bestaat dat dit kostentechnisch interessant is in vergelijking met het gebruik van Databricks.
- Apache Airflow is onafhankelijk van een serviceprovider. Dit betekent dat deze oplossing (met enige aanpassingen) kan functioneren binnen zowel een AWS- als bijvoorbeeld een Azure-omgeving. Dit zorgt dus voor een flexibele oplossing voor de toekomst.

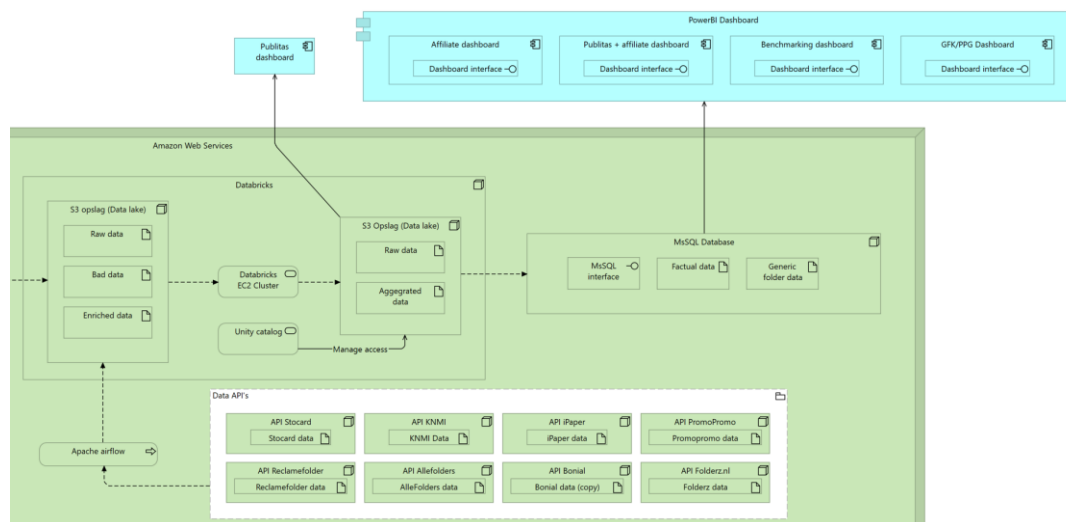
In samenvatting is de keuze dus gemaakt op basis van drie belangrijke aspecten: interne expertise, kostenefficiëntie en flexibiliteit.

Omdat de keuze gebaseerd is op de expertise van één persoon, kan de vraag gesteld worden: wat als de data-engineer vertrekt? Dit risico wordt op twee manieren opgevangen. Publitas is een organisatie die geen kantoor heeft naast het kantoor van WePublish en iedereen verder dus thuis

werkt. Publitas hanteert daarom de policy dat alles nauwkeurig gedocumenteerd dient te worden binnen Notion. Notion is een omgeving waarin alle informatie van de organisatie is opgeslagen, inclusief de handleidingen. Verder is Apache Airflow Python based en mag ervan uitgegaan worden dat een data-engineer hiermee overweg kan, aangezien dit een van de meest gebruikte taal is bij het werken met data.

Vanuit deze S3-opslag worden de gegevens doorgestuurd naar twee onderdelen. Het eerste onderdeel is het bekende Publitas-dashboard, waar alleen de gegevens naartoe worden gestuurd die afkomstig zijn van de eigen kanalen van de retailer, oftewel de gegevens die worden getrackt doormiddel van de Snowplow-datatracker. Daarnaast worden de geaggregeerde gegevens ook doorgestuurd naar de MsSQL-server. Deze gegevens worden gebruikt om de volgende dashboards van data te voorzien: Affiliate-dashboard, Publitas + affiliate dashboard, Benchmarking-dashboard en het GFK/PPG-dashboard.

In onderstaande afbeelding zijn de zojuist benoemde onderdelen te zien: de opslag van de data binnen Databricks, de API-verzameling die via Apache Airflow wordt opgebouwd en opgehaald en tot slot de dashboards van Publitas en WePublish.



Figuur 14 IT-architectuur API Data & MsSQL

Externe data en standaardisatie

Eén van de problemen die WePublish momenteel ondervindt bij de datavergaring en datakwaliteit van de externe kanalen, is dat de data van externe kanalen zoals de folderplatformen niet gestandaardiseerd is. Elk platform hanteert zijn eigen methode om data te versturen/klaar te zetten, wat varieert tussen Excel-sheets, data scraping (Wat is data scraping? Definitie en gebruik | Okta, sd) of een API (Marc Lohuis, persoonlijke communicatie, 15 februari 2024). Daarnaast geeft de technisch directeur aan dat er grote verschillen zijn in hoe zij data aanleveren, sommige folderplatformen leveren data op dag niveau, anderen op week niveau en soms wordt er gewerkt met geaggregeerde data in plaats van ruwe data (volledige interviewtranscript is opgenomen in Bijlage 1).

Om dit in de toekomst te verbeteren en correct te implementeren, is standaardisatie vereist. Dit moet op twee niveaus worden gerealiseerd: ten eerste een uniforme manier van het aanleveren van data vanuit de folderplatformen en ten tweede moeten er eisen worden gesteld aan de data die door de folderplatformen worden aangeleverd.

De technisch directeur geeft aan dat folderplatformen als Reclamefolder, Folderz.nl en AlleFolders hiervoor open staan en mee willen denken. Verder zullen de overige platformen en internationale

platformen in de toekomst niet willen achterblijven. Hier zal ook een verantwoordelijkheid van de retailer komen te liggen. Want zij hebben hier groot belang bij dat de data op een goede manier op één centrale punt wordt gevisualiseerd, zij dienen deze wens dus ook bij de folderplatformen uit te spreken.

Huidige en toekomstige externe data dashboarding

Als er naar de huidige filtering wordt gekeken binnen PowerBI, is dit niet overzichtelijk. Omdat retailers (vaak) geen foldernaam meesturen, maakt WePublish deze zelf aan op basis van een weeknummer. Daarnaast zie je in figuur 15 een voorbeeld waar meerdere folders met hetzelfde weeknummer worden getoond. Dit is kortom niet overzichtelijk en zorgt voor verwarring.

- ☐ Gamma week 04 2024 weekenddeal - WePublish
- ☐ GAMMA Week 4 2024 - Alle Folders
- ☐ Gamma week 04 2024 weekenddeal
- ☐ GAMMA week 4 2024 - Folders.nl
- ☐ Gamma Week 4 2024 - Reclamefolder
- ☐ GAMMA week 4 2024 - WePublish

Figuur 15 PowerBI folder filters

Om dit in de toekomstige situatie te verhelpen, is het volgende aan te raden, WePublish verzorgt de verspreiding van de folder op folderplatformen, zoals Reclamefolder, AlleFolders en Stocard. Bij het verspreiden moet in de toekomst een identifier worden meegestuurd. Deze identifier moet in het WePublish-systeem gekoppeld zijn aan één unieke folder.

Wanneer het folderplatform vervolgens de folderdata aanbiedt volgens de gekozen oplossing van WePublish, moet deze identifier worden meegegeven. Dit zorgt ervoor dat de folderplatformen geen naamgevingen hoeven toe te passen en dat WePublish de data kan koppelen onder de generieke folderdata van henzelf.

Wanneer deze data vervolgens wordt gevisualiseerd binnen het dashboard van WePublish, zal er in het folderfilter één folder zichtbaar zijn waar alle data onder gekoppeld staat. Vervolgens is deze data filterbaar door het filter Platform toe te voegen, in dit filter zullen de verschillende platformen zichtbaar zijn en zo kunnen de inzichten op platformniveau getoond worden.

Feitentabel

In de feitentabel hoeven geen grote aanpassingen te worden doorgevoerd, omdat de kanalen WePublish en Folders.nl al goed functioneren. Het is echter belangrijk om, zoals hierboven genoemd, dit te standaardiseren door van de externe folderplatformen te eisen dat zij de folder identifier meesturen en de benodigde data ter beschikking stellen die voldoet aan de eisen van Publitas/WePublish. Zie figuur 16 voor de oude en nieuwe feitentabel van WePublish. Verder wordt in de onderstaande tabel per kolom een korte uitleg gegeven.

Factual tabel (Oud)	Factual tabel (Nieuw)
View_id	Folder_identifier
Date	View_id
Retailer_id	Date
Variant_id	Retailer_id
Page_id	Page_id
Interaction_id	Interaction_id
Page_view_time	Page_view_time
Event_action	Event_action
Total_events	Total_events

Figuur 16 Feitentabel oud en nieuw

Event naam	Beschrijving
Folder_identifier	Folder_identifier is het unieke ID die meegestuurd wordt met de folder tijdens de verspreiding en die de retailer dient terug te sturen met de data.
View_id	View_id bevat het platform waarvan de data afkomstig is.
Date	Datumveld
Retailer_id	Retailer_id is de unieke identifier van een retailer.
Page_id	Page_id is de unieke identifier van een digitale folder pagina.
Interaction_id	Interaction_id is de unieke identifier van een interactie in de digitale folder.
Page_view_time	Page_view_time is een numerieke waarde waar de leestijd wordt meegegeven.
Event_action	Event_action toont aan over wat voor type event het gaat. Voorbeelden zijn: unique_view, page_view en click
Total_events	Total_Events zijn de opgesomde events.

Tabel 2 Beschrijving feitentabel

STAR-model

Voor de toekomstige situatie is een STAR-model opgesteld waarin de Folder_identifier is opgenomen. Dit model is opgesteld voor de dashboards Affiliate Dashboard en Affiliate + Publitas dashboard. Dit model is gebaseerd op de benodigde dimensie data en feiten data. Metingen die binnen PowerBI bestaan of in de toekomst nodig zijn om het dashboard te realiseren, zijn niet meegenomen in dit model. Hiervoor is gekozen om het model 'schoon' en overzichtelijk te houden. WePublish kan dit model aanhouden als basis in de toekomst. In figuur 17 is het STAR-model opgenomen.



Figuur 17 STAR-model

Onderhoud en Ondersteuning:

Als de toekomstige IT-architectuur is gerealiseerd door Publitas/WePublish, moet deze ook worden onderhouden. Dit zal worden uitgevoerd door het DevOps-team van Publitas, dat verantwoordelijk is voor alles met betrekking tot IT-architectuur. Het DevOps-team bestaat uit de volgende medewerkers: Gabor Vincze (DevOps lead), Umer Khan (DevOps engineer) en Vibor Jain (Site Reliability Engineer).

Qua infrastructuur hoeven er geen grote aanpassingen te worden gemaakt. Het DevOps-team krijgt echter de volgende onderdelen toegevoegd die het moet onderhouden: MsSQL-server, API's, dashboards en de data pipelines gemaakt met Apache Airflow.

Verder moet er worden gekeken naar de workload van de huidige situatie, omdat er meer data zal doorstromen vanuit de nieuwe kanalen. Uit het interview met Umer blijkt echter dat zij ruimte hebben ingeregeld voor een hogere workload dan momenteel nodig is. Als dit niet voldoende blijkt te zijn, zal hier tijdig op worden gereageerd door middel van monitoring en het versturen van meldingen (volledige interviewtranscript is opgenomen in Bijlage 6)

4.4.2 Conclusie deelvraag 4

“Hoe dient de toekomstige IT-architectuur eruit te zien voor Publitas, waarin zij de interne en externe folderdata kunnen opslaan?”

De IT-architecturen van WePublish en Publitas dienen samengevoegd te worden in de AWS-omgeving van Publitas. Hiervoor is onder andere voor gekozen omdat is de migratie van de WePublish-klanten van de FMP back-end naar de Publitas back-end al in gang is gezet. Dus zullen alle consumenten in de toekomst gebruikmaken van de Publitas viewer, waarbij de interne folderdata tracking plaatsvindt binnen Snowplow en de gegevens worden opgeslagen binnen Databricks.

Verder dient de externe folder data via een ETL-proces, gebouwd met Apache Airflow, opgehaald te worden en in de Databricks-omgeving geplaatst te worden. Deze externe data wordt overigens gestandaardiseerd naar de eisen van Publitas. Dit wordt gedaan via een Folder_identifier die wordt meegestuurd tijdens de verspreiding. Tot slot wordt de WePublish-tak van de organisatie verantwoordelijk voor de affiliate dashboard en de affiliate + Publitas dashboards. Met andere woorden, de verdiepende dashboards. Deze dashboards worden in de volgende deelvraag uitgewerkt.

4.5 Deelvraag 5

In dit hoofdstuk worden de resultaten besproken van de volgende deelvraag: “Hoe kan de geconsolideerde data effectief worden gevisualiseerd om waardevolle inzichten te bieden aan de stakeholders?”

4.5.1 Resultaten / Functioneel ontwerp

Om antwoord te geven op deelvraag 5 is een functioneel ontwerp opgesteld. Het doel van het functioneel ontwerp is om Publitas en WePublish een leidraad te bieden over hoe zij interne en externe data kunnen visualiseren binnen het toekomstige dashboard. Dit functioneel ontwerp is specifiek gericht op de visualisatie van zowel interne als externe data in de geavanceerde dashboards en niet op het in-app dashboard van Publitas.

Het doel is niet om een volledig nieuw dashboard te ontwikkelen, maar om een eerste opzet te maken die een goed beeld geeft van het toekomstige dashboard. De elementen die de belanghebbenden als positief beoordelen in het huidige dashboard zullen indien nodig behouden blijven, terwijl de verbeterpunten zullen worden aangepakt. Naast de primaire belanghebbenden wordt ook aandacht besteed aan de literatuur van Stephan Few.

In het functioneel ontwerp worden de volgende onderwerpen behandeld:

- Requirements en prioritering volgens de MoSCoW-methode.
- Beschrijving van de functionaliteiten van het dashboard.
- Wireframes, onderbouwd door literatuur en rekening houdend met de primaire stakeholders.

4.5.1.1 Requirements analyse

In dit hoofdstuk worden de functionele en niet-functionele requirements (Functional and Nonfunctional Requirements: Specification and Types, 2023) beschreven, gecategoriseerd en geprioriteerd aan de hand van de MoSCoW methode. De requirements zijn in kaart gebracht door middel van verschillende expertinterviews en mijn eigen observaties vanuit mijn rol als data-analist bij WePublish.

In de requirementsanalyse worden de vereisten opgesteld voor het dashboard 'Publitas + folderplatformen'. Hiervoor is gekozen omdat het 'affiliate dashboard' op dezelfde principes wordt gebouwd en hiervoor geen volledig nieuwe requirementsanalyse nodig is.

Stakeholders

Uit de stakeholderanalyse is gebleken dat de volgende drie de primaire stakeholders zijn, namelijk de business consultants, sales medewerkers en het management, waarvan de business consultants en sales medewerkers de sleutelfiguren zijn. Om de requirements op te stellen zijn twee interviews gehouden. Het interview met de sales/accountmanager is terug te vinden in bijlage 11 en ook is er een interview afgenomen met het management, in dit geval de technisch directeur en verantwoordelijke voor de tooling van WePublish. Dit interview is terug te vinden in bijlage 1. Tot slot is met de business consultant een workshop gehouden om de opgesteld requirements te valideren en aan te vullen.

Requirements & categorisatie

Op basis van de twee interviews, eigen kennis en observaties als Data-analist binnen WePublish zijn twee tabellen opgesteld: één voor de functionele requirements en één voor de niet-functionele requirements. In deze tabellen zijn de requirements benoemd, is het type toegewezen en bij elke requirement is een uitleg gegeven over wat het precies inhoudt.

Beide primaire stakeholders geven aan al tevreden te zijn met het huidige dashboard en zien dit als een goed uitgangspunt. Waar extra waarde gecreëerd kan worden, is het uitbreiden van het data aanbod door meerdere folderplatformen te koppelen, het Publitas-thema hierin te verwerken, bugs te verhelpen en de gebruikersvriendelijkheid te verbeteren.

Functionele requirements	Type	Toelichting
Gebruikersauthenticatie	Business	Het systeem moet het mogelijk maken dat de gebruiker kan inloggen met een toegewezen gebruikersnaam en wachtwoord.
Actualiteit data	User	De data die de gebruiker ziet in het dashboard moet actueel zijn. Dit dient doormiddel van een dagelijkse sync geüpdatet te worden.
Error en waarschuwing	Business	Wanneer de dagelijkse sync niet succesvol is dient DevOps hier een melding van te krijgen via e-mail.
Integratie externe gegevensbronnen	User	De gebruikers dienen via het dashboard data van externe gegevensbronnen zoals folderplatformen in te kunnen zien.
Data filtering	User	De gebruikers moeten data kunnen filteren om gerichte inzichten zichtbaar te maken. Dit dient op platform-, datum- en folderniveau gedaan te worden.
Data exports	User	Visualisaties dienen data exports te ondersteunen om de onderliggende data naar bijvoorbeeld Excel te exporteren.
Interactieve visualisaties	User	Visualisaties dienen interactief samen te werken. Voorbeeld: Wanneer een gebruiker één week selecteert in een grafiek dienen KPI Buttons alleen data van deze specifieke week te tonen.
Overkoepelende en gerichte inzichten	User	Gebruikers dienen data op overkoepelend niveau kunnen inzien van alle gekoppelde kanalen. Met overkoepelend wordt bedoeld op de KPI's folder impressies, folder views en clicks op dag-, week-, maand- en jaarniveau. Verder dienen ze ook data op folder en folderpagina niveau te zien.
Jaar vs Jaar	User	Gebruikers dienen de mogelijkheid te hebben om de KPI's folder impressies, folder views en clicks op week niveau met elkaar te vergelijken.
Folder overzicht	User	Gebruikers dienen een overzicht te hebben van de belangrijkste KPI's in één tabel/overzicht.

Tabel 3 Functionele requirements

Niet functionele requirements	Type	Toelichting
Dashboard prestaties	User	Het dashboard moet binnen 5 seconden reageren bij het laden van het dashboard en het aanpassen van filters.
Schaalbaarheid	Business	Het dashboard moet effectief omgaan met een groeiende dataset en de hoeveelheid gebruikers.
Gebruikersvriendelijk	User	Het dashboard dient gebruikersvriendelijk te zijn, ook voor gebruikers zonder technische achtergrond.
Beschikbaarheid	Business	Het dashboard dient een uptime van minimaal 99% hebben om de klanttevredenheid te waarborgen.

Onderhoud dashboard	Business	Onderhoud dient gemakkelijk uitgevoerd te kunnen worden, wat verder geen effect heeft op de live omgeving.
Dataprivacy	Business	Retailer data dient goed afgeschermd te worden van elkaar door middel van Row Level Security.
Publitas huisstijl	Business	Publitas wil consistentie in branding om de merkidentiteit te versterken. Daarom dient de huisstijl van Publitas toegepast te worden.
Documentatie	Business	Het dashboard moet uitgebreid gedocumenteerd worden, zowel op gebruikersniveau als op beheerdersniveau, om de gebruikersvriendelijkheid te bevorderen.

Tabel 4 niet functionele requirements

4.5.1.2 Prioritering MoSCoW

In de tabel hieronder zijn de requirements geprioriteerd aan de hand van de MoSCoW-methode.

Voor Publitas zijn een aantal aspecten van cruciaal belang (must have) voor het dashboard. Dit omvat de dataveiligheid en privacy, de prestaties en beschikbaarheid en het tonen van actuele, correcte inzichten van interne en externe gegevensbronnen. Daarnaast is dit een tool vanuit Publitas en dient dit zichtbaar te zijn door middel van de huisstijl.

De should have's zijn ingesteld op aspecten die extra waarde kunnen leveren aan de gebruiker. Dus, welke datapunten hebben toegevoegde waarde, hoe gebruikersvriendelijk is het dashboard en wat betreft het onderhoud ervan.

Tot slot hebben we de could have's. Dit zijn voor Publitas de aspecten die niet direct nieuwe waarde toevoegen, maar gericht zijn op het gebruikersgemak. Denk hierbij aan het exporteren van data, overzichten en documentatie. Zonder deze punten is het dashboard echter nog steeds werkbaar.

Requirement	Type requirement	MoSCoW
Gebruikersauthenticatie	Functioneel	Must have
Error en waarschuwing	Functioneel	Must have
Integratie externe gegevensbronnen	Functioneel	Must have
Data filtering	Functioneel	Must have
Overkoepelende en gerichte inzichten	Functioneel	Must have
Dashboard prestaties	Niet functioneel	Must have
Schaalbaarheid	Niet functioneel	Must have
Beschikbaarheid	Niet functioneel	Must have
Dataprivacy	Niet functioneel	Must have
Publitas huisstijl	Niet functioneel	Must have
Actualiteit data	Functioneel	Should have
Interactieve visualisaties	Functioneel	Should have
Jaar vs Jaar	Functioneel	Should have
Gebruikersvriendelijk	Niet functioneel	Should have
Onderhoud dashboard	Niet functioneel	Should have
Data exports	Functioneel	Could have

Folder overzicht	Functioneel	Could have
Documentatie	Niet functioneel	Could have

Tabel 5 MoSCoW prioritering

Validatie requirements

Tijdens het opstellen van de requirements is gebruikgemaakt van informatie uit twee interviews en kennis uit eigen observaties als data-analist. Om de requirements te valideren, is een workshop gehouden met een business consultant. Tijdens deze workshop is gekeken naar de volledigheid van de requirements en hun prioritering. Op basis hiervan zijn de tabellen verder aangescherpt en gevalideerd.

Het opstellen van de requirements en het validatieproces is uitgevoerd met vier personen. De breedte van dit proces is mede beïnvloed doordat er bij Publitas nog onvoldoende kennis is over het gebruik van data en het ondersteunen van de retail, waardoor hun kennis niet kon worden benut. Verdere details over de interne en externe stakeholders zijn beschreven in de laatste alinea van de discussie.

4.5.1.3 Kritische prestatie indicatoren digitale folder

In dit hoofdstuk worden de belangrijkste KPI's van de digitale folder vermeld en uitgelegd. Deze KPI's zijn opgesteld op basis van eigen kennis die is opgedaan in vier jaar werkervaring als data-analist en als verantwoordelijke voor het maken van data-analyses voor zowel interne als externe stakeholders.

Folder impressie: Een folder impressie is een KPI die alleen gemeten wordt op folderplatformen. Dit event wordt gemeten wanneer een bezoeker bijvoorbeeld Folders.nl bezoekt en de foldercover van de folder op het platform ziet staan. De consument hoeft hiervoor de foldercover dus niet aan te klikken.

Folder view: Een folder view is de KPI die wordt gemeten wanneer een bezoeker de folder aanklikt en de folder dus opent.

Open click-through rate (Open CTR%): De open click-through rate is het percentage van bezoekers dat vanuit een folder impressie de folder aanklikt en dus een folder view genereert.

Click: Een click wordt gemeten wanneer een bezoeker in de folder een hyperlink aanklikt.

Click-through rate (CTR%): De click-through rate is het percentage van folder views waarbij er wordt doorgeklikt in de folder. Dit wordt berekend door het aantal clicks te delen door het aantal folderweergaven.

Gem. Gelezen pagina's: De gemiddelde gelezen pagina's is het gemiddelde aantal pagina's dat één bezoeker leest per folder view.

Gem. Leestijd: De gemiddelde leestijd is het gemiddelde aantal seconden dat een bezoeker op één pagina doorbrengt.

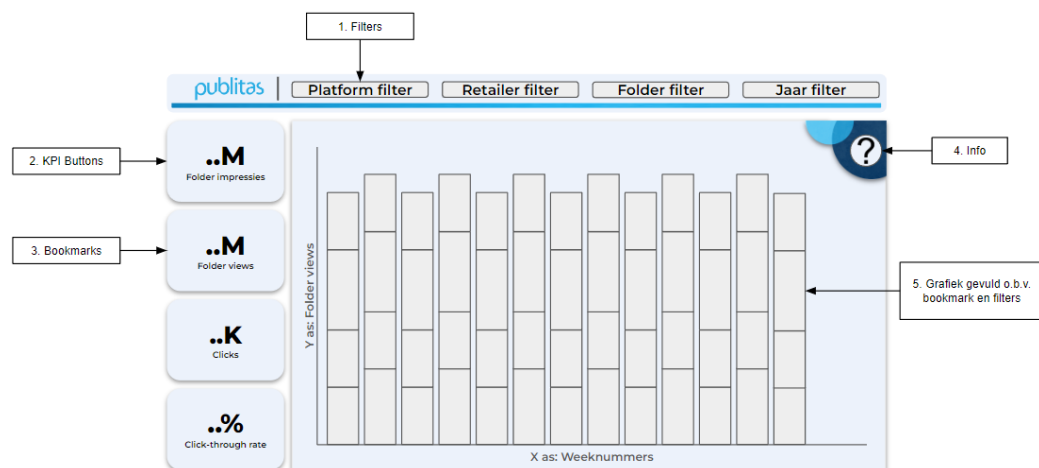
4.5.1.4 Wireframes affiliate + Publitas dashboard

In dit hoofdstuk worden de verschillende wireframes van het toekomstige dashboard toegelicht. Het is als volgt opgebouwd: Eerst staat de titel, gevolgd door de uitleg en onder de uitleg staat het wireframe. Tijdens de beschrijving is er rekening gehouden met de literatuur van Stephen Few. Deze literatuur is verder toegelicht in hoofdstuk 2.2.6.

Overzicht KPI's & Kanalen

De stakeholders kunnen met het dashboard "Overzicht KPI's en kanalen" de belangrijkste KPI's inzien. De belangrijkste KPI's zijn de folderimpressies op folderplatformen, dus hoe vaak de cover wordt gezien. Folder views, het aantal keren dat de folder wordt gelezen. Clicks, het aantal keer dat er in de folder wordt geklikt en tot slot de Click-through rate, het percentage van de bezoekers dat klikt in de folder. Uit vier jaar werkervaring als Data-analist blijkt dat dit de belangrijkste KPI's zijn voor zowel interne als externe stakeholders.

De KPI's worden weergegeven aan de hand van KPI-buttons aan de linkerkant, waarbij rekening is gehouden met kleurgebruik. Er is een lichte achtergrondkleur gebruikt met een krachtige kleur voor de data. Deze KPI-buttons zijn klikbaar, waardoor het dashboard meebeweegt en de geselecteerde KPI in de grafiek wordt weergegeven. In de grafiek zal in de toekomst de data van de verschillende platformen worden getoond met de huisstijlkleur van het platform, zoals geel voor Folders.nl en lichtblauw voor Publitas. Bovendien zijn aan de bovenzijde filters opgenomen en is geprobeerd zo min mogelijk niet-data-pixels te gebruiken naast de pixels voor de huisstijl.

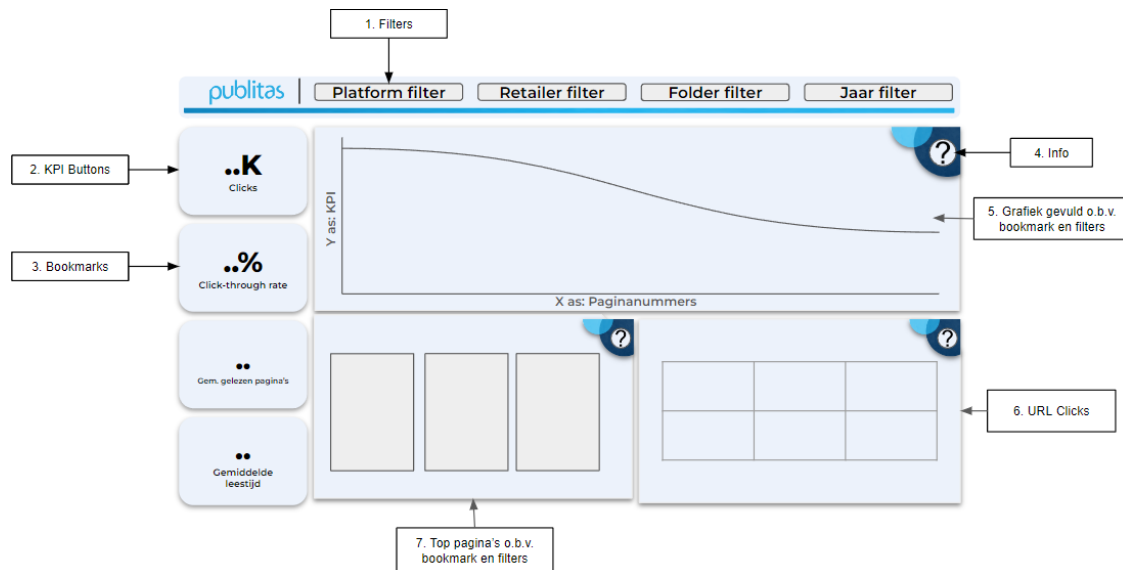


Figuur 18 Wireframe KPI's & kanalen

Folder inzichten

Met het dashboard "Folderinzichten" kunnen stakeholders dieper inzoomen op het niveau van individuele folders. Aan de linkerkant zijn de belangrijkste KPI's opgenomen die van belang zijn wanneer er op folder niveau wordt gekeken. Dit zijn de volgende KPI's: Clicks, dus het aantal keer dat er in de folder is geklikt, Click-through rate, het percentage van het totale aantal bezoekers dat klikt, gemiddeld aantal gelezen pagina's en gemiddelde leestijd in seconden.

Deze KPI-buttons zijn op dezelfde wijze opgesteld als de vorige, waarbij rekening is gehouden met kleurgebruik en ze zijn ook klikbaar. De bovenste grafiek toont op paginaniveau de geselecteerde KPI. Bijvoorbeeld, voor clicks worden de clicks per pagina weergegeven. Linksonder wordt de top 3 weergegeven van de geselecteerde filters en KPI. Tot slot, rechtsonderin, worden de clicks per URL getoond, gekoppeld aan het paginanummer.

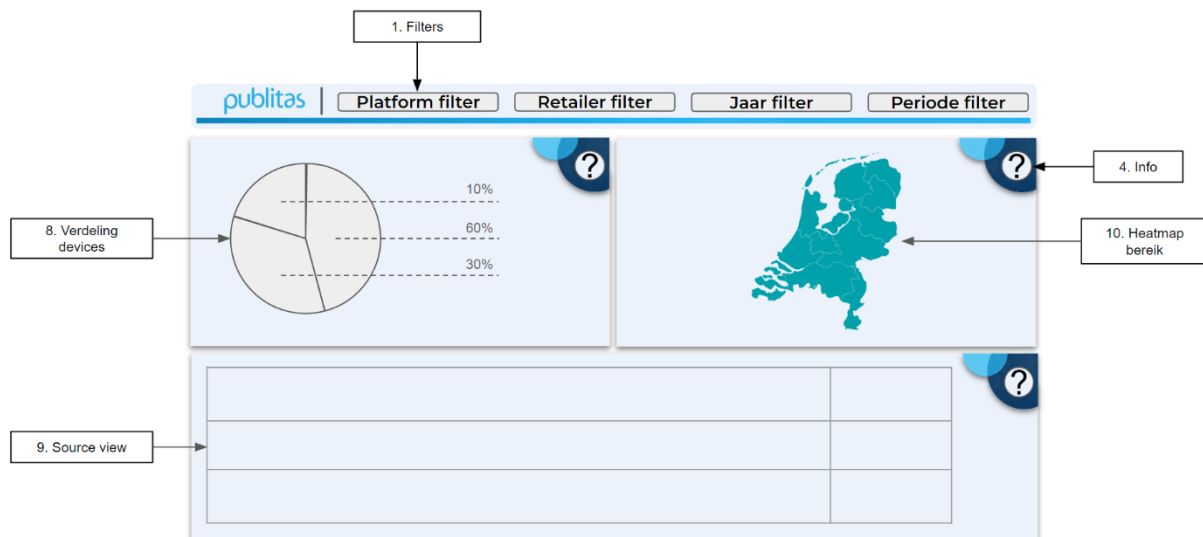


Figuur 19 Wireframe folder inzichten

Kanalen en source

Met het dashboard "Kanalen en bronnen" kan de stakeholder inzien waar de folderviews vandaan komen, verdeeld over drie aspecten: de gebruikte apparaten (welk type apparaat het meest wordt gebruikt), het kanaal dat verantwoordelijk is voor het bereik van de folder (bijvoorbeeld vanuit een advertentie op Nu.nl of Facebook, of standaard vanaf de eigen website) en de oorsprong van het bereik.

Bij het ontwerpen van het dashboard is, net als bij het bovenstaande dashboard, rekening gehouden met de huisstijl en het gebruik van pixels. Verder is hier gekozen voor een cirkeldiagram omdat er altijd maar 3 verschillende mogelijkheden zijn en dit niet varieert.

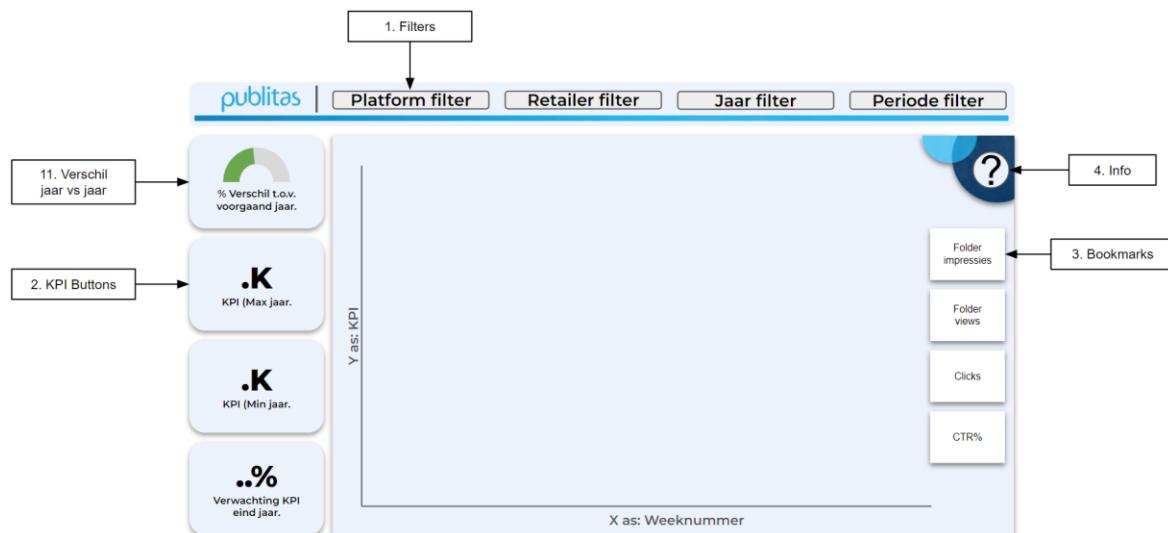


Figuur 20 Wireframe kanalen en source

Jaar vergelijking

Met het dashboard "Jaarvergelijking" kan de stakeholder de belangrijkste KPI's op jaar-/weekniveau met elkaar vergelijken. Door rechts een KPI te selecteren en via het jaarfilter twee verschillende jaren te kiezen, zal dit in de grafiek in het midden worden geplot. Hier kan de stakeholder vervolgens aflezen hoe zij het doen, zowel aan de hand van de grafiek als de KPI-buttons aan de linkerkant. Ook wordt het percentage verschil getoond. Verder kan aan de hand van het periode filter een specifieke range worden geselecteerd om specifieke periodes met elkaar te vergelijken.

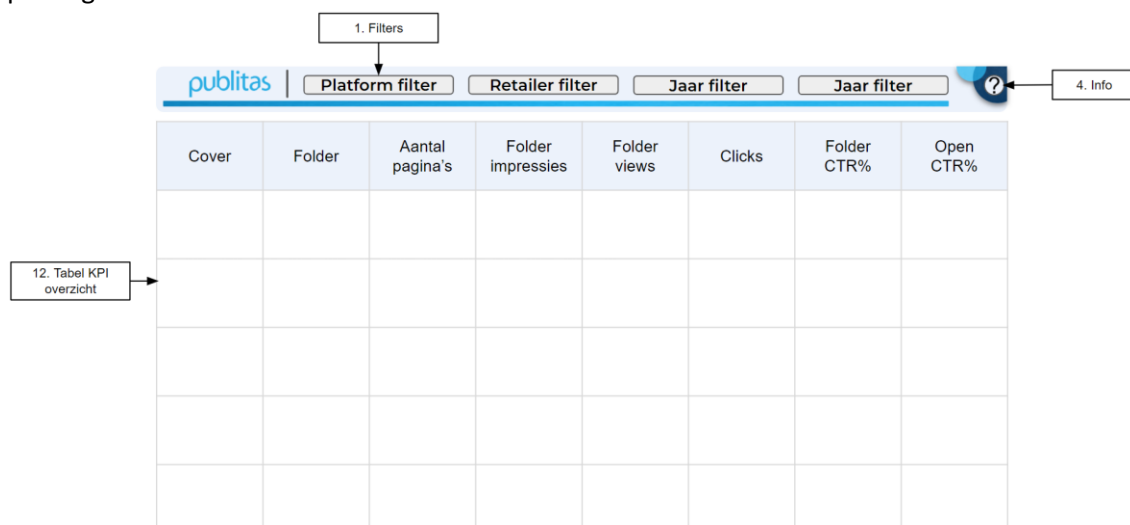
Bij het ontwerpen van het dashboard is, net als bij het bovenstaande dashboard, rekening gehouden met de huisstijl en het gebruik van pixels.



Figuur 21 Wireframe jaar vergelijking

Folder overzicht

Met het dashboard "Folder overzicht" kan de stakeholder op een overzichtelijke manier de belangrijkste KPI's inzien van de folders van de verschillende platformen. Deze data wordt in dit dashboard weergegeven in een tabel die de stakeholders gemakkelijk kunnen exporteren naar Excel. Verder is er rekening gehouden met de huisstijl van Publitas en zijn er geen overbodige niet data pixels gebruikt.



Figuur 22 Wireframe folder overzicht

Toelichting informatiebalken

In de figuren van de opgestelde wireframes zijn nummers opgenomen die de onderdelen van de wireframes aangeven. In dit hoofdstuk worden deze aspecten van de wireframes kort en bondig uitgelegd

- 1) **Filters:** Dit zijn de filters die de gebruiker van het dashboard kan instellen. Deze filters tonen alleen data die is toegewezen aan de gebruiker. Dus, bijvoorbeeld, zal Gall & Gall alleen de platformen en folders zien waar zij toegang toe hebben. Verder zijn dimensiefilters zoals jaar opgenomen, evenals het filter voor de retailer. Het retailerfilter is opgenomen omdat er gebruikers zijn met toegang tot meerdere kanalen, zoals Albert Heijn en Albert Heijn België.
- 2) **KPI Buttons:** Dit zijn buttons die een specifieke KPI laten zien. Deze buttons bewegen mee met de filters en geselecteerde data in de grafieken.
- 3) **Bookmarks:** De KPI Buttons zullen ook als Bookmark fungeren. Dit houdt in dat er een klikbaar vak over de buttons wordt geplaatst. Wanneer dit vak wordt aangeklikt wijzigen de grafieken naar visualisaties die te maken hebben met de geselecteerde KPI.
- 4) **Info:** Dit is een hover-over button die informatie toont over het onderdeel waar het bijstaat.
- 5) **Grafiek gevuld o.b.v. bookmark en filters:** Dit toont de geselecteerde data visueel in een staaf- of lijngrafiek. Welke data wordt getoond, is afhankelijk van de ingestelde filters en bookmarks.
- 6) **URL Clicks filters:** Dit is een tabel die de URL, clicks en paginanummer toont. Dit geeft de gebruiker inzicht in welke URL's goed scoren.
- 7) **Toppagina's o.b.v. bookmark en filters:** Dit toont de beste pagina's. Hier wordt data getoond op basis van de geselecteerde filters en de bookmark.
- 8) **Verdeling devices:** In deze grafiek kan de gebruiker aflezen vanuit welk type device het bereik komt. Hier heb je drie typen devices: tablet, mobiel en desktop.
- 9) **Source view:** Hier kan de gebruiker aflezen vanuit welke source het bereik afkomstig is. Voorbeelden hiervan zijn: GAMMA.nl/folder of Facebook.nl.
- 10) **Heatmap bereik:** Hier kan de gebruiker aflezen vanuit welke locatie de bezoeker afkomstig is, rekening houdend met de AVG-wetgeving. Dit is voornamelijk interessant voor internationale folders.
- 11) **Verschil jaar vs jaar:** Hier kan de gebruiker aflezen of zij het goed doen of slecht doen ten opzichte van het voorgaande jaar.
- 12) **Tabel KPI-overzicht:** Dit is een tabel waarin KPI's per folder worden getoond.

4.1.5 Conclusie deelvraag 5

“Hoe kan de geconsolideerde data effectief worden gevisualiseerd om waardevolle inzichten te bieden aan de stakeholders?”

Uit het functioneel ontwerp kan worden geconcludeerd dat Publitas, door rekening te houden met de opgestelde requirements en de literatuur van Stephan Few, een effectief dashboard kan opzetten dat zowel interne als externe data effectief visualiseert. De opgestelde wireframes bieden Publitas een leidraad bij het visualiseren van de data, waarbij rekening wordt gehouden met de gestelde requirements en de literatuur. Deze wireframes zijn opgesteld zonder een specifieke visualisatietool te vermelden. De visualisaties bevatten geen 'speciaal' ontwikkelde aspecten en zijn dus te visualiseren in verschillende visualisatietools.

5 Conclusie

In dit hoofdstuk worden de deelconclusies kort toegelicht, waarna er antwoord wordt gegeven op de hoofdvraag *“Hoe kan Publitas, na de overname van WePublish, op een effectieve manier de data toepassingen van beide partijen consolideren en wegzetten in een datawarehouse omgeving, om dit vervolgens te visualiseren.”*

Uit deelvraag 1 *“Wat zijn de consequenties van de fusie voor de informatievoorziening strategie met betrekking tot folderdata van Publitas en welk advies is er voor de toekomstige strategie omtrent de informatievoorziening?”* komt een tweeledig advies. Publitas dient zich te richten op haar eigen platformen/websites van de retailers. Het is echter van belang dat zij gebruikmaken van de expertise die beschikbaar is bij WePublish rondom data en optimalisatie van de digitale folders. De WePublish-tak dient zich daarnaast te oriënteren op de breedte van het data-aanbod. Hierbij wordt bedoeld op affiliate platformen en het bouwen van koppelingen tussen deze platformen en de data-visualisatie en dashboards ervan.

Als er vervolgens in deelvraag 2 *“Hoe ziet de huidige data-architectuur eruit bij zowel WePublish als Publitas en wat zijn de sterke punten en zwakke punten ervan?”* gekeken wordt naar de huidige IT-architecturen, zien we dat WePublish een omgeving heeft binnen Azure en Publitas binnen AWS. WePublish schiet momenteel tekort in het waarborgen van de datakwaliteit en schaalbaarheid. De beveiliging is echter wel goed ingeregeld. Bij de IT-architectuur van Publitas zijn deze zaken goed geregeld en worden ze beheerd door het DevOps-team.

In deelvraag 3 *“Hoe dient de geconsolideerde data van WePublish en Publitas vormgegeven te worden?”* is vervolgens gekeken naar de getrackte data uit beide folder viewers en of dit aansluit bij de vraag van de stakeholders. Hieruit blijkt dat de tracking en gemodelleerde data van Publitas niet voldoet aan de wensen van de primaire stakeholders. Dit is bij WePublish wel het geval, Publitas dient dit dus aan te passen naar de wensen van de stakeholders. De historische data van Publitas is echter niet verloren en kan worden samengevoegd met de WePublish-data. Dit vereist echter wel enige aanpassingen.

In deelvraag 4 *“Hoe dient de toekomstige IT-architectuur eruit te zien voor Publitas, waarin zij de interne en externe folderdata kunnen opslaan?”* is een technisch ontwerp opgezet waarbij een toekomstig ontwerp wordt getoond voor de interne en externe data binnen de AWS-omgeving van Publitas. Hier is onder andere voor AWS gekozen omdat de organisatie al heeft besloten om de WePublish-klanten naar de Publitas back-end te migreren. Verder is in het technisch ontwerp een affiliate plan uitgewerkt, waarbij Publitas op gestandaardiseerde wijze de data kan ophalen en visualiseren binnen het affiliate dashboard. Voor dit zogenoemde dashboard is uit de stakeholderanalyse, requirementsanalyse en de literatuur van Stephen Few gebleken hoe dit vormgegeven moet worden. In het functioneel ontwerp uit deelvraag 5 *“Hoe kan de geconsolideerde data effectief worden gevisualiseerd om waardevolle inzichten te bieden aan de stakeholders?”* zijn deze resultaten opgenomen.

Als er vervolgens naar de hoofdvraag wordt gekeken, kan op basis van de deelconclusies geconcludeerd worden dat Publitas zich moet blijven richten op haar eigen platform, waarbij WePublish zich dient te concentreren op de affiliate platformen. De IT-architecturen van beide organisaties dienen te worden geconsolideerd binnen de AWS-omgeving, waar ze interne en externe data samenbrengen. Deze data kan vervolgens effectief worden gevisualiseerd aan de hand van de opgestelde wireframes, waarbij rekening wordt gehouden met de belangen van de primaire stakeholders. De wireframes hebben als toekomstig doel de sales te bevorderen en ondersteuning te bieden aan de data-consultancy.

6 Discussie

Tijdens het uitvoeren van het onderzoek is de validiteit en betrouwbaarheid van het onderzoek gewaarborgd door zoveel mogelijk triangulatie toe te passen. Dit betekent dat er vanuit meerdere invalshoeken naar de onderwerpen wordt gekeken, in hoofdstuk 3 zijn de verschillende methoden per deelvraag toegelicht. Verder zijn de resultaten waar nodig getoetst bij de belanghebbenden.

Het feit dat ik het onderzoek uitvoer en tegelijkertijd werk als data-analist bij WePublish, kan invloed hebben gehad op mijn objectiviteit. Ik heb geprobeerd dit zoveel mogelijk te reguleren door nauwlettend naar de opdracht te kijken en waar mogelijk informatie van beide organisaties in te winnen.

Verder komen de resultaten enigszins overeen met mijn verwachtingen. Ik merkte bij het opzetten van het onderzoek dat er al een duidelijke voorkeur was voor consolidatie binnen de IT-architectuur van Publitas. Daarnaast bleek uit het onderzoek dat dit ook de meest passende keuze was. Ook zag ik een sterke splitsing tussen het development en datateam van beide organisaties en een verschillende blik van het management op deze onderdelen van de organisatie. Hier had ik verwacht dat er een wens zou ontstaan om dit samen te voegen. Dit is echter niet het geval en het advies luidt dan ook om beide zich te richten op één specifiek onderdeel. Dus interne data bij Publitas en externe data bij WePublish, beide onderdelen zullen wel in dezelfde IT-architectuur worden opgenomen.

Tijdens het onderzoek is een enquête opgesteld en verzonden naar de customer success managers van Publitas, dit is alleen naar de Publitas customer succes managers gestuurd omdat WePublish deze niet heeft. Uit meerdere enquêtevragen kwam vervolgens sterk naar voren dat er bij Publitas nog te weinig met folderdata gedaan wordt. Hoewel deze data aanwezig is, is de communicatie naar retailers op dit gebied zeer terughoudend in vergelijking met WePublish. Om deze reden werden deze resultaten niet als waardevol beschouwd voor het onderzoek en is de informatie niet opgenomen. Desalniettemin heeft dit mij wel een beeld gegeven van de huidige stand van zaken.

Verder begon ik het onderzoek met één beperking in deelvraag 4. Deelvraag 4 ging als volgt: "Hoe dient de toekomstige datawarehousing-oplossing binnen Databricks eruit te zien voor Publitas, waarin zij de geconsolideerde data kunnen opslaan?". Hier zat dus een afbakening ingebouwd rondom Databricks. Uit het onderzoek bleek dat dit breder getrokken moest worden omdat Databricks niet de gehele IT-architectuur representeert en daarom is de deelvraag aangepast om dit te realiseren. De vernieuwde deelvraag is: "Hoe dient de toekomstige IT-architectuur eruit te zien voor Publitas, waarin zij de interne en externe folderdata kunnen opslaan?"

Voor het opstellen van de requirements en wireframes voor de toekomstige dashboardoplossing in deelvraag 5 is gekeken naar de primaire stakeholders, deze stakeholders zijn interne stakeholders. Naast deze interne stakeholders zijn er ook externe stakeholders die belang hebben bij het dashboard. Tijdens het onderzoek is alleen rekening gehouden met de primaire stakeholders, dus uitsluitend de interne stakeholders. De keuze hiervoor is gemaakt omdat het dashboard als primair doel heeft de sales en data-consultancy te bevorderen. Om het dashboard extra inhoud te geven, is het goed verstandig ook de secundaire externe stakeholders in de toekomst hierbij te betrekken. Als dit niet wordt gedaan, bestaat de kans dat de secundaire stakeholders het dashboard niet als meerwaarde zien, waardoor niet alle mogelijke waarde uit het dashboard wordt gehaald.

7 Aanbevelingen

In dit hoofdstuk worden de verschillende aanbevelingen opgesteld en toegelicht. Deze aanbevelingen zijn toekomstgericht en gelden voor de organisatie Publitas. Per alinea wordt één aanbeveling toegelicht.

Uit het onderzoek blijkt dat WePublish al actief bezig is met folderdata. Ze ondersteunen retailers bij folderoptimalisaties en integreren folderdata in de afdelingen sales en consultancy voor commerciële doeleinden. Bij Publitas is uit het onderzoek gebleken dat dit minder het geval is. Echter, vanuit het management wordt hier wel een meerwaarde in gezien. Tijdens het uitvoeren van de enquête onder de customer success managers werd echter duidelijk dat zij nog weinig gebruikmaken van folderdata en zich voornamelijk richten op de tevredenheid en functies van de tools van Publitas. Deze aanbeveling is gebaseerd op deze bevindingen: er dient meer kennisdeling plaats te vinden omtrent data om de customer success managers een beter inzicht te bieden in de meerwaarde van data en de kansen die dit met zich meebrengt.

In de discussie is al benoemd dat bij het ontwerp van de wireframes alleen rekening is gehouden met de interne stakeholders. De wireframes voldoen aan hun verwachtingen en dus aan de verwachtingen van de primaire stakeholders van het project. Echter zijn de interne stakeholders niet de enigen die het dashboard gaan gebruiken. Daarom wordt de aanbeveling gedaan dat Publitas vervolgonderzoek moet uitvoeren bij de externe retailers die het dashboard zullen gebruiken. Deze resultaten dienen vervolgens geverifieerd te worden met de interne stakeholders om zo tot een definitief ontwerp te komen.

Als er vervolgens wordt gekeken naar de ambitie van de WePublish-tak, gericht op externe folderdata van affiliate platformen, blijkt dat ze zich sterk willen positioneren als dé data-partij. Dit betekent dat de Retail wereld bij Publitas/WePublish moet zijn voor een totaalbeeld en diepgaande inzichten in folderdata. Dit is momenteel een goed streven en haalbaar binnen een redelijk tijdsbestek. De aanbeveling is echter dat ze verder moeten kijken dan alleen het tonen van data-inzichten via een dashboard. Hier ligt de mogelijkheid om meer te doen met de beschikbare data, waarbij machine learning en AI interessante onderwerpen zijn. Een concreet voorbeeld is het trainen van een machine learning-model met data, zodat het een optimalisatiescore kan genereren voor een folder. Deze score geeft aan hoe goed een folder is geoptimaliseerd op basis van metrics zoals lengte van de folder, aantal links, spreiding van producten, enzovoort. Kortom, Publitas moet niet stilstaan nadat alle data gevisualiseerd is, maar zou verder moeten gaan met het verkennen van geavanceerde en innovatieve mogelijkheden.

Tot slot heb ik één algemene aanbeveling op basis van mijn observatie en het uitvoeren van het onderzoek. Tijdens het onderzoek merkte ik op dat Publitas en WePublish beide een eigen visie hebben op het gebied van data. Uit het onderzoek blijkt dat dit goede keuzes zijn, echter gaat er veel expertise en kennis verloren als deze twee losstaande facetten blijven binnen de organisatie. Daarom is mijn aanbeveling om het onder één management te brengen, met bijvoorbeeld één eindverantwoordelijke en in ieder geval wekelijkse meetings om de samenwerking en kennisdeling te bevorderen.

Bibliografie

- (A.), A.-J. (2023, april 24). *Waarom de papierprijzen flink zijn gestegen*. Opgehaald van janssenpers: <https://www.janssenpers.nl/2022/06/13/waarom-de-papierprijzen-flink-zijn-gestegen/>
- Addison, T. (2022, februari 18). *Grocery leaflets going digital? - Hammer Agency*. Opgehaald van Hammer Agency: <https://hammeragency.eu/grocery-leaflets-going-digital/>
- Azap, B. (2022, Augustes 29). *Wat is schaalbaarheid?* Opgehaald van phoenixNAP IT Glossary: <https://phoenixnap.nl/woordenlijst/schaalbaarheid>
- Cluster Management - Amazon Elastic Container Service*. (sd). Opgehaald van aws.amazon: <https://docs.aws.amazon.com/AmazonECS/latest/developerguide/clusters-concepts.html>
- consilium.europa. (2023, juni 06). *Markteffecten van de Russische invasie in Oekraïne: EU-respons*. Opgehaald van consilium europa: <https://www.consilium.europa.eu/nl/policies/eu-response-ukraine-invasion/impact-of-russia-s-invasion-of-ukraine-on-the-markets-eu-response/#:~:text=De%20energieprijzen%20zijn%20sinds%20medio,over%20de%20energiev%20oorzaking%20is%20ontstaan>
- Data governance | Uitleg en definitie - ICTinformatiecentrum.nl*. (2023, November 08). Opgehaald van ICTinformatiecentrum: <https://www.ictinformatiecentrum.nl/data-management/data-governance>
- Few, S. (2006). Information Dashboard Design. In S. Few, *Information Dashboard Design* (p. 74). O'Reilly.
- Few, S. (2006). Information Dashboard Design. In S. Few, *Information Dashboard design* (pp. 84-85). O'Reilly.
- Few, S. (2006). Information Dashboard Design. In S. Few, *Information Dashboard Design* (p. 67). O'Reilly.
- Few, S. (2006). Information Dashboard Design. In S. Few, *Information Dashboard Design* (p. 82). O'Reilly.
- Few, S. (2006). Information Dashboard Design. In S. Few, *Information Dashboard Design* (p. 97). O'Reilly.
- Functional and Nonfunctional Requirements: Specification and Types*. (2023, november 30). Opgehaald van altexsoft: <https://www.altexsoft.com/blog/functional-and-non-functional-requirements-specification-and-types/>
- Genau. (2023, januari 23). *Semigestructureerde of half-gestructureerde interviews in je scriptie*. Opgehaald van Scribbr: <https://www.scribbr.nl/onderzoeksmethoden/semigestructureerd-interview/>
- Gertjanschop, G. (2020, april 30). *MoSCoW - ontwerpcriteria voor oplossingen - Managementmodellensite*. Opgehaald van Managementmodellensite: <https://managementmodellensite.nl/moscow-ontwerpcriteria-voor-oplossingen/>
- Gertjanschop, G. (2023, Februari 04). *GAP-analyse vanuit Ist naar Soll - Managementmodellensite*. Opgeroepen op November 23, 2023, van Managementmodellensite: <https://managementmodellensite.nl/gap-analyse-vanuit-ist-naar-soll/>
- Gertjanschop, G. (2023, 11 28). *SWOT - Managementmodellensite*. Opgehaald van Managementmodellensite: <https://managementmodellensite.nl/swot-analyse/>

- GfK. (2022, 11 23). *Stevige opmars van de digitale reclamefolder*. Opgehaald van <https://www.gfk.com/>: <https://www.gfk.com/press/Opmars-digitale-reclamefolder>
- ICTInformatiecentrum. (2023, November 8). *Datakwaliteit | uitleg en definitie - ICTInformatiecentrum.nl*. Opgehaald van ICTInformatiecentrum: <https://www.ictinformatiecentrum.nl/data-management/datakwaliteit>
- Markteffecten van de Russische invasie in Oekraïne: EU-respons*. (2023, juni 06). Opgehaald van consilium europa: <https://www.consilium.europa.eu/nl/policies/eu-response-ukraine-invasion/impact-of-russia-s-invasion-of-ukraine-on-the-markets-eu-response/#energy>
- MoSCoW – projectmanagementsite*. (sd). Opgehaald van projectmanagementsite: <https://projectmanagementsite.nl/moscow/>
- S3 Loader | Snowplow Documentation*. (2023, September 21). Opgehaald van docs.snowplow: <https://docs.snowplow.io/docs/pipeline-components-and-applications/loaders-storage-targets/s3-loader/>
- Stakeholdersanalyse – projectmanagementsite*. (sd). Opgeroepen op 02 08, 2024, van <https://projectmanagementsite.nl/stakeholdersanalyse/>
- Succes begint met een Brainstormsessie - Samen 100+ ideeën uitdenken?* (2023, januari 26). Opgehaald van IMPRES: <https://www.impres.nl/strategie-en-advies/brainstormsessie/#:~:text=Wat%20is%20brainstorming%3F,antwoorden%20verzamel%20voor%20jouw%20vraagstuk.>
- Talend, T. (sd). *What is data mapping? Definition and examples*. Opgehaald van Talend - A Leader in Data Integration & Data Integrity: <https://www.talend.com/resources/data-mapping/#:~:text=Data%20mapping%20is%20the%20process,it%20accessible%20to%20decision%20makers.>
- The Brussels Times*. (2022, oktober 08). Opgehaald van brusselstimes: <https://www.brusselstimes.com/301833/supermarket-promotions-leaflets-continue-to-thrive-in-the-digital-age>
- Vleeshouwers. (2023, Oktober 30). *Hoe gebruik je de gap-analyse in je scriptie?* Opgehaald van Scribbr: <https://www.scribbr.nl/modellen/gap-model-servqual/>
- Wat betekent schaalbaarheid voor systemen en services?* (sd). Opgeroepen op Februari 07, 2024, van Lucidchart: <https://www.lucidchart.com/blog/nl/wat-is-schaalbaarheid#:~:text=Een%20schaalbaar%20systeem%20is%20een,aan%20de%20vraag%20te%20voldoen.>
- Wat is data scraping? Definitie en gebruik | Okta*. (sd). Opgeroepen op 02 15, 2024, van Okta, Inc.: <https://www.okta.com/nl/identity-101/data-scraping/#:~:text=Bij%20data%20scraping%20wordt%20informatie,analyses%2C%20verwerking%20of%20een%20>
- What is Airflow™? — Airflow Documentation*. (sd). Opgehaald van Airflow.apache: <https://airflow.apache.org/docs/apache-airflow/stable/index.html>

Bijlage 1 – Interview Marc Lohuis

Interviewer: Nick Reintjes

Geïnterviewde: Marc Lohuis

Functie geïnterviewde: Technisch directeur

Onderwerp 1: Strategie & Visie folderdata

Nick: voor mijn onderzoek ben ik bezig met de huidige informatievoorziening strategie met betrekking tot folder data in kaart te brengen van WePublish en Publitas. Ik heb jou gevraagd voor dit interview omdat jij verantwoordelijk bent voor alles omtrent de techniek van WePublish. Om te beginnen wil ik je vragen wat jouw huidige functie is en wat jij dagelijks doet.

Marc: Jouw baas zijn. Verder houd ik me bezig met alle zaken rondom techniek. Dit kan hands-on zijn als ik een afdeling moet bijspringen, maar ik houd me voornamelijk bezig met de langetermijnplanning van zowel product, data en development. Waaronder dus ook de strategie rondom folder data.

Nick: Wat is de huidige strategie binnen de organisatie met betrekking tot data- en informatievoorziening, specifiek gericht op folder data?

Marc: Dit bestaat uit 3 onderdelen. Ten eerste dient het voor intern gebruik, zodat we kunnen monitoren hoeveel verkeer we genereren en daarmee de schaalbaarheid kunnen optimaliseren. Ten tweede is het bedoeld om prestaties inzichtelijk te maken voor de retailer, zoals welke folders en campagnes goed presteren. Tot slot en naar mijn mening het belangrijkste, wordt het gebruikt voor het vergelijken tussen verschillende platformen. Dit omvat folderplatformen zoals Reclamefolder en AlleFolders, evenals het eigen platform van retailers, zoals GAMMA.nl. Het voordeel hiervan is dat we al een voet tussen de deur hebben bij de retailer en ze meer kunnen bieden dan alleen data van eigen platform. We hebben hier in de Nederlandse markt al veel succes mee en nu zijn we aan het kijken of we dit kunnen uitbreiden in bijvoorbeeld de Duitse markt.

Nick: Wat zijn de consequenties voor de huidige strategie met betrekking tot folder data van WePublish, nu Publitas is gefuseerd met WePublish?

Marc: Gui, de baas van Publitas, wilde het folderplatform Folders.nl aanvankelijk verkopen omdat hij de meerwaarde niet zag. Maar nu ziet hij toch in hoe waardevol het is vanwege de vele kansen bij retailers en op de markt. Gui is zelfs zo enthousiast dat hij het folderplatform Promobutler in België heeft gekocht. En nu kijken we zelfs of we meer samenwerkingen kunnen opzetten met andere folderplatformen en die aan ons dashboard kunnen koppelen. We ervaren dus geen negatieve effecten, maar voornamelijk een positieve impuls.

Nick: Wat is vanuit het perspectief van WePublish de visie met betrekking tot folder data?

Marc: Als we ons richten op onze visie met betrekking tot data en dan voornamelijk folder data, zal het dashboard van de data van de eigen platformen, zoals GAMMA.nl, voornamelijk plaatsvinden in het Publitas-dashboard. Bij WePublish zullen we ons concentreren op meer diepgaande dashboards, zoals affiliate dashboards voor de folderplatformen als Alle Folders, Reclamefolder en buitenlandse platformen in de toekomst. Deze gegevens worden ook gekoppeld aan de eigen platformen in die landen, waarbij we de informatie ophalen vanuit het Publitas-dashboard. Met onze omvang en toonaangevende klanten geloof ik dat we aanzienlijke eisen kunnen stellen aan folderplatformen. We kunnen bijvoorbeeld aangeven welke datapunten we nodig hebben en dat platformen die hier niet aan voldoen, niet worden opgenomen in onze dashboards en folder verspreiding.

Nick: Wat zijn de consequenties voor de huidige visie met betrekking tot folder data van WePublish, nu Publitas is gefuseerd met WePublish?

Marc: Minimale consequenties, ik denk dat dit al goed doorgesproken is en dat zij hier ook de meerwaarde van zien. Maar we moeten wel alert zijn op de ontwikkelingen in de foldermarkt, aangezien dit een krimpende markt is voor onze eigen platformen. Steeds meer Nederlandse retailers gebruiken alternatieve methoden via verschillende mediakanalen om consumenten te bereiken en dus minder investeren in digitale folders. We zien nu al retailers die ermee stoppen. Ik denk dat het essentieel is om ons te onderscheiden in de markt door middel van data, waardoor we een toonaangevende organisatie kunnen worden. In de toekomst zullen functies binnen de online folder waarschijnlijk niet meer voldoende zijn.

Nick: Is er op dit moment weerstand vanuit Publitas met betrekking tot de huidige strategie en visie?

Marc: Ik zou het niet per se weerstand noemen, maar ze zijn wel zeer voorzichtig. Het kost veel moeite om ze te overtuigen. Op dit moment is er een tweedeling. Alles wat betrekking heeft op onze eigen platformen zal in de toekomst worden ontwikkeld door Publitas. Wat betreft verspreiding richting andere platformen en het dashboard zal de verantwoordelijkheid bij het team van WePublish in Oldenzaal liggen.

Nick: Wat zijn volgens jou de risico's wanneer WePublish zich niet richt op de huidige visie?

Marc: Ik denk dat de folder markt krimpt. Nieuwe retailers brengen niet veel folders uit en traditionele retailers verminderen terwijl e-commerce groeit. Om te groeien, moeten we ons meer richten op markten waar ons marktaandeel klein is. Bijvoorbeeld, in Duitsland zijn we vier keer kleiner dan in Nederland, terwijl het een veel grotere markt is.

Onderwerp 2: SWOT-Analyse

Nick: Wat zijn volgens jou de Sterktes van WePublish

Marc:

- Sterke naamsbekendheid/merknaam in Nederland, België en Duitsland.
- Uitgebreide kennis over het optimaliseren van folders.
- expertise op het gebied van de foldermarkt in de Benelux.

Nick: Wat zijn de zwaktes van WePublish

Marc: De tooling en IT-architectuur is nog niet goed samengevoegd tussen de organisaties.

Nick: Wat zijn volgens jou de kansen voor WePublish

Marc:

- Faciliteren van de verspreiding van folderplatformen nationaal en internationaal.
- Beschikbaar budget van de retailer op korte termijn, door de kosten van papier en verschuiving van de markt naar digitaal.

Nick: Wat zijn volgens jou de bedreigingen voor WePublish

Marc: 1 grote bedreiging namelijk, namelijk concurrentie van alternatieve e-commerce en digitale marketingkanalen.

Nick: Duidelijk, bedankt in ieder geval voor je tijd Marc. Mocht ik nog vragen hebben kom ik bij je terug!

Addendum interview: Requirements deelvraag 5

Nick: Ha Marc, fijn dat je nog even tijd hebt voor een kort vervolg van het interview. Voor deelvraag 5 ga ik het eerste ontwerp van het dashboard opstellen dat gebaseerd is op de voorgaande deelvragen. Om tot een goed ontwerp te komen, dienen eerst de requirements te worden opgesteld. Hierover heb ik nog een aantal vragen, omdat jij als technisch directeur en eindverantwoordelijke voor het dashboard een primaire stakeholder bent.

Marc: Ja helemaal goed, vertel.

Nick: Als je naar het dashboard kijkt, wat zijn dan de minimale functionele eisen die het moet hebben?

Marc: Ik wil graag een paar punten aanhalen: de authenticatie moet goed zijn, zodat retailers alleen toegang hebben tot hun eigen gegevens en een eigen account hebben waarmee ze kunnen inloggen in het dashboard. Daarnaast is het belangrijk dat de data actueel is en dat de verschillende folderplatformen gekoppeld zijn.

Nick: Authenticatie en folder platformen zijn duidelijk, je zegt verder actueel. Is dit real-time, dagelijks wekelijks?

Marc: Actueel doel ik op dagelijks. Dus een dagelijkse sync. Dit omdat je altijd één complete dag wil hebben om te kijken naar de prestaties. De real-time data is beschikbaar in de database mocht dit gewenst zijn.

Verder wil ik nog toevoegen dat er in de toekomst een goede documentatie moet zijn van het dashboard en misschien deels ook opgenomen in de dashboarding zelf om de gebruiker van het dashboard meer ondersteuning te bieden. We zijn nu nog te veel tijd kwijt aan vragen die we eerder zouden kunnen opvangen.

Nick: Als we specifieker kijken naar het dashboard, zoals de opmaak, filtering, visualisaties enzovoort, zijn er dan nog specifieke eisen?

Marc: Hiervoor kun je het beste Sales, Ibe benaderen of zelf beantwoorden. Jullie zitten namelijk dicht bij de retail en hebben een beter inzicht hierin.

Nick: Duidelijk en nogmaals bedankt. Ik denk hier al een goede start mee te maken. Mochten er nog punten te binnen schieten kom ik nog even bij je terug.

Marc: Yes, geen probleem.

Bijlage 2 – Interview Paul Raya

Interviewer: Nick Reintjes

Geïnterviewde: Paul Raya

Functie Geïnterviewde: Data lead Publitas

Onderwerp 1: Strategie & Visie folderdata

Nick: For my research, I am in the process of mapping out the current information/data strategy related to folder data at WePublish and Publitas. I have asked you for this interview because you are responsible for all things related to the technology of Publitas. To start, I would like to ask what your current role is and what you do on a daily basis.

Paul: Hello Nick, sounds good. My name is Paul, and I am the Team Lead for Data and CRM. This includes the strategy as well, although it is still being discussed with others ofcourse.

Nick: What is the current strategy within the organization regarding data, specifically focused on folder data?

Paul: Yes, the strategy revolves around placing the data in such a way that there is one truth. Currently, we are doing this in DataBricks. We do this for both internal data and external folder data.

Nick: And what is the main purpose of the external folder data?

Paul: The primary goal of the folder data is to further assist the retailer. Currently, we provide them with KPI's. The dashboard we use for this purpose, however, is relatively new.

Nick: What are the consequences for the current strategy concerning folder data from Publitas, now that Publitas has merged with WePublish?

Paul: The strategy for assisting the retailer through data has not changed. However, there is a need to consider how we want to proceed with the development of the dashboard. Since there is more expertise available on the WePublish side, we need to work towards a single solution.

Nick: I talked to Marc today, and he thinks we need a single solution for retailer data dashboards. He sees Publitas handling this, and WePublish focusing more on the folder platforms. What is your take on this?

Paul: That sounds good; however, we will need to discuss this further. This does not change the fact that we still need the expertise from WePublish.

Nick: From the perspective of Publitas, what is the vision regarding folder data?

Paul: The vision aligns quite well with the strategy. We aim to provide the retailer with a single source of truth based on data. In addition to this vision, we aim to use machine learning in the future to predict which folders will perform well and which ones may not. We intend to support these predictions with data.

Nick: Is there currently any resistance from Publitas/WePublish regarding the current strategy and vision?

Paul: No.

Nick: In your opinion, what are the risks if Publitas does not align with the current vision?

Paul: I think this poses a significant risk. In the current society and business landscape, there is a growing demand for data, which is often used to validate decisions. If we cannot provide this, I see a considerable risk that the customer will seek out a party that can meet this demand.

Nick: And what is needed to realize the strategy/vision in the future?

Paul: time, expertise, and a architecture that facilitates it.

Onderwerp 2: SWOT-Analyse

Nick: According to you, what are the strengths of Publitas?

Paul:

- Strong brand name
- Our tooling
- Active in many international markets

Nick: What are the weaknesses of Publitas?

Paul: Publitas and WePublish are not yet aligned in terms of tooling and architecture

Nick: According to you, what are the opportunities for Publitas?

Paul:

- Further develop dashboarding
- Predictive analysis.

Nick: According to you, what are the threats to Publitas?

Paul: Rapidly growing demand for data insights and substitute marketing products.

Nick: Thanks Paul, if I have any further questions, I will reach out to you!

Bijlage 3 – 5 Krachten model van porter

5 Krachten model van Porter

Om verder in het onderzoek een duidelijk beeld te krijgen van de toekomstige strategie met betrekking tot de data- en informatievoorziening van Publitas, is de concurrentie geanalyseerd met behulp van de methode het 5 Krachten model van Porter. In deze analyse is de concurrentiekracht getoetst aan de hand van de vijf krachten.

Vergaring data

Om het Vijfkrachtenmodel van Porter in te vullen, is data verzameld uit verschillende bronnen. In de eerste plaats zijn gegevens verzameld via interviews, welke zijn opgenomen in bijlage 1 & 2.

Daarnaast zijn deze gegevens verder aangevuld door middel van eigen observaties. Het aangevulde model is vervolgens gevalideerd en verder aangevuld tijdens een workshop met Marc Lohuis, Ramon Eerdman en Ibe Hundling. Het doel van deze workshop was om het Vijfkrachtenmodel te valideren en af te ronden.

Grafische weergave & uitwerking 5 krachten model van porter



Figuur 23 5 Krachten model porter

Bedreiging van nieuwe toetreders

Ieder punt van de kracht "Bedreiging van nieuwe toetreders" zal hier worden toegelicht.

- De ontwikkelingskosten voor zowel de software om de digitale folders te verwerken als de software om ze op een specifieke locatie te plaatsen, zijn aanzienlijk hoog.
- WePublish en Publitas hebben een zeer goede naam die gelinkt is aan kwaliteit. Hoe sterk dit is, kan verschillen per land of regio. Maar over het algemeen is de naam positief.
- Het begrijpen van wat goed werkt in de digitale folder en hoe dit varieert per branche vereist ervaring en data. Het optimaliseren van digitale folders heeft verder geen standaardoplossing en vereist altijd een frisse blik.

- Doordat Publitas een belangrijke speler is in de digitale folderwereld, profiteert het van schaalvoordelen. Software en andere functies worden ontwikkeld voor een meerdere klanten en kunnen daardoor eerder worden gerechtvaardigd.

Onderhandelingsmacht van leveranciers

Ieder punt van de kracht "Onderhandelingsmacht van leveranciers" zal hier worden beschreven.

- Publitas is afhankelijk van grote dominante leveranciers van cloud en database toepassingen.

Onderhandelingsmacht van afnemers

Ieder punt van de kracht "Onderhandelingsmacht van afnemers" zal hier worden beschreven.

- Publitas werkt voor zeer grote retailers, zoals Albert Heijn en Action in de Benelux. Aangezien deze retailers een aanzienlijk groot deel van de omzet vertegenwoordigen, hebben ze aanzienlijke onderhandelingsmacht.
- De retailers die klant zijn van Publitas stellen altijd hoge eisen aan de kwaliteit. Dit doen zij omdat de digitale folder een belangrijk marketingkanaal is. Daarom zijn de eisen voor klanttevredenheid hoog.
- Retailers werken samen met mediabureaus die de inkoop doen voor meerdere bedrijven of partijen. Omdat dit een groep is, neemt hun onderhandelingsmacht toe, aangezien het risico om omzet te verliezen aanzienlijk stijgt.

Bedreiging van substituten

Ieder punt van de kracht "Bedreiging van substituten" zal hier kort worden beschreven.

Digitale marketingkanalen kunnen worden beschouwd als substituten voor de digitale folder. De volgende substituten vormen een directe bedreiging voor Publitas.

- Goed ontworpen webshops.
- Social Media
- E-mail nieuwsbrieven
- Mobiele apps

Intensiteit van de concurrentie

Ieder punt van de kracht "Intensiteit van de concurrentie" zal hier kort worden beschreven.

- Prijsconcurrentie is altijd relevant. Als er meer aanbieders van de specifieke software op de markt zijn, zal er sterke concurrentie ontstaan en bestaat het risico dat klanten Publitas verlaten of minder willen betalen voor dezelfde functionaliteiten.
- Publitas biedt weinig maatwerk. Voordat een functie wordt ontwikkeld voor een retailer, moet het zeker zijn dat meerdere retailers deze zullen gebruiken, zodat de functie gerechtvaardigd is. Dit kan eventuele deuren openen voor concurrentie.
- Wanneer de concurrentie zichtbaarder wordt via reclame en andere mediakanalen, kan dit leiden tot een grotere maat van concurrentie.

Analyse en conclusie

Per kracht wordt aangegeven hoe groot de mate van bedreiging is en waarom dit het geval is.

Bedreiging van Nieuwe Toetreders:

Laag. De dreiging van nieuwe toetreders is laag vanwege de zeer hoge ontwikkelingskosten van de software voor het verwerken en online zetten van de digitale folders, schaalbaarheid, sterke merknaam en de expertise rondom digitale folders.

Onderhandelingsmacht van Leveranciers:

Laag. Publitas is afhankelijk van grote leveranciers op het gebied van cloud- en databasetoepassingen, echter blijkt uit de historie dat deze partijen geen buitensporige prijsstijgingen doorvoeren of onredelijke eisen stellen.

Onderhandelingsmacht van Afnemers:

Hoog. De onderhandelingsmacht van de afnemers is hoog. Publitas heeft enkele grote retailers en media partijen met een aanzienlijk aandeel in de omzet. Deze partijen hebben naast het hoge aandeel ook een hoge kwaliteitseis. Kortom deze afnemers hebben een aanzienlijke macht.

Bedreiging van Substituten:

Gemiddeld. De dreiging van substituten is altijd aanwezig, zeker in deze tijd is het gevaar aanwezig als er gekeken wordt naar digitale marketingkanalen waarin een retailer in zou kunnen investeren. De bedreiging scoort gemiddeld omdat retailers nog altijd erg gehecht zijn aan zowel de papieren als digitale folder, wat Publitas nu wel ziet is dat er een verschuiving plaats vindt richting digitaal. Dit is positief.

Intensiteit van Concurrentie:

Gemiddeld. Hetgeen waar de concurrentie voornamelijk sterk in is, is maatwerk. Publitas moet als grote organisatie eerst kunnen rechtvaardigen waarom zij voor één specifieke klant bijvoorbeeld een feature ontwikkelen. Naast dat de concurrentie dus maatwerk oplossing aanbiedt, bieden zij dit ook aan voor een scherpe prijs.

Conclusie

De concurrentiedruk scoort laag tot gemiddeld. De dreiging van nieuwe toetreders tot de digitale foldermarkt en de machtspositie van leveranciers in dit gebied worden als laag beoordeeld en vereisen op dit moment geen extra aandacht.

De substitutiekraacht en concurrentie intensiteit scoren gemiddeld. Het is belangrijk dat Publitas hier rekening mee houdt en dit actief gaat monitoren. Daarnaast zijn de grote klanten en media partijen van groot belang en vormen dus een reëel risico. Het wegvallen van deze retailers zou aanzienlijke omzetsdalingen veroorzaken.

Het is voor Publitas cruciaal om de tevredenheid van de retailers die verantwoordelijk zijn voor een groot deel van de omzet te waarborgen. Dit kan bijvoorbeeld worden bereikt door extra aandacht en het leveren van interessante extra data-inzichten.

Bijlage 4 – SWOT-Analyse

Voor Publitas is een SWOT-analyse opgesteld die moet bijdragen aan het formuleren van een strategisch advies omtrent de data/informatie voorziening omtrent folderdata. Dit wordt gedaan op basis van de geïdentificeerde kansen, bedreigingen, sterktes en zwaktes die voortkomen uit de SWOT-analyse.

Vergaring data

Om de SWOT-analyse in te vullen, is data verzameld via verschillende methodes. De gebruikte methoden hiervoor zijn interviews, die te vinden zijn in de volgende bijlagen 1 & 2. Daarnaast is er ook gebruikgemaakt van input uit het opgestelde 5 krachten model van Porter, dit model is terug te vinden in bijlage 3

Grafische weergave & SWOT Analyse



Figuur 24 SWOT-analyse

Sterktes

Ieder punt van de “Sterktes” zal hier kort worden toegelicht:

- Zowel WePublish, als Publitas hebben sterke merknamen in de wereld van digitale folders.
- Zowel WePublish als Publitas hebben hun eigen tools ontwikkeld, wat als voordeel heeft dat zij niet afhankelijk zijn van externe partijen op dat gebied.
- Doordat Publitas actief is op verschillende markten over de gehele wereld, beschikt Publitas over internationale kennis met betrekking tot digitale folders, die zij kunnen toepassen in de verschillende regio's.
- Gedurende de jaren hebben zowel WePublish als Publitas veel kennis opgedaan met betrekking tot het optimaliseren van digitale folders en de data dat dit ter weeg brengt.

Zwaktes

Ieder punt van de “Sterktes” zal hier kort worden toegelicht:

- Als gevolg van de fusie is het klantenbestand van Publitas momenteel versnipperd wat betreft de productie tooling die Publitas gebruikt. Een deel van digitale folder wordt verwerkt via de WePublish-tools, terwijl een ander deel via de Publitas-tools. Deze versnippering vormt een zwakte omdat de processen nog niet naadloos op elkaar aansluiten, wat kan leiden tot waardevermindering voor de retailers.
- Ditzelfde geldt voor de IT-architectuur. Momenteel draaien er 2 IT-architecturen die nog niet gefuseerd kunnen worden omdat dit nog niet naadloos op elkaar overloopt. Zolang dit het geval is zal een retailer die overgezet wordt naar de Publitas tooling waardeverliezen.
- Door de fusie verlopen de overige processen nog niet naadloos op elkaar en verschillen de werkwijzen.

Kansen

Ieder punt van de “Sterktes” zal hier kort worden toegelicht:

- Momenteel is er een verschuiving gaande van papier naar digitaal. Bovendien zijn er veel retailers die uitsluitend nog online folders publiceren en geen gebruik meer maken van papieren folders, twee voorbeelden hiervan zijn Bol.com en Holland & Barrett. Kortom, hier komt budget vrij.
- Qua data-consultancy loopt de Benelux voor op bijvoorbeeld Zuid-Amerika en Spanje. Nu Publitas een aanzienlijke hoeveelheid kennis heeft overgenomen door de fusie met WePublish, kunnen zij deze kennis gaan inzetten om de overige markten naar een hoger niveau te tillen.
- Momenteel wordt er nog te sterk geleund op het evalueren van prestaties uit het verleden. Echter, de kans ligt in de toekomst, doormiddel van machine learning en predictive analyses kan Publitas de retailer helpen met bijvoorbeeld het vullen van de digitale folder.
- WePublish heeft ervaring in het verspreiden van folders voor Nederlandse retailers. Door deze expertise internationaal toe te passen, versterkt Publitas haar positie op de markt. Ook kan Publitas samen met toonaangevende retailers bij de verspreidingsplatforms data opvragen, waardoor het een belangrijke speler wordt voor inzichten in digitale folderdata.

Bedreigingen

Ieder punt van de “Sterktes” zal hier kort worden toegelicht:

- De belangrijkste externe bedreigingen op dit moment zijn de substituten. Hoewel Publitas beschikt over een sterk klantenbestand en uitgebreide expertise op het gebied van folders, constateren ze dat veel retailers ook gebruikmaken van andere digitale marketingmiddelen. Deze vormen dus een reëel gevaar.

Analyse en conclusie

In dit hoofdstuk worden de sterkten en zwaktes geanalyseerd, evenals de kansen en bedreigingen op basis van het voorgaande hoofdstuk. Ten slotte wordt er een conclusie aan verbonden.

Analyse sterkten en zwakten

Op basis van de informatie beschreven in voorgaand hoofdstuk heeft Publitas de volgende sterkten. Het beschikt over een sterke merknaam, eigen ontwikkelde tools en aanzienlijke kennis met betrekking tot de digitale foldermarkt in zowel de Benelux als internationaal.

Desondanks zijn er enkele zwakke punten die voornamelijk voortkomen uit de fusie met WePublish. Dit resulteert in een versnipperde set tools, twee verschillende IT-architecturen die nog niet zijn geconsolideerd en te veel processen die verschillen tussen de twee organisaties.

Analyse kansen en bedreigingen

Als we vervolgens kijken naar de kansen, zien we het volgende: de digitale folder is in opmars, mede doordat papier duurder wordt en er meer aandacht wordt besteed aan duurzaamheid. Kortom, er komt hiervoor meer budget vrij. Verder is het helpen van de retailer met het vullen van de folder door middel van predictive analyses een reële kans voor Publitas. Bovendien kan Publitas datgene wat WePublish al doet, namelijk de verspreiding van folders naar folderplatformen, internationaal gaan inzetten.

Tegenover deze kansen staat echter één duidelijke bedreiging. Dat zijn substituten; in de digitale wereld zijn er talloze mogelijkheden als het gaat om digitale marketing, zoals sociale media en nieuwsbrieven. Dit vormt dus een reële bedreiging.

Marc gaf verder aan dat hij hier en daar al retailers hoort praten over het verder inzetten van andere digitale marketingmiddelen en acht het een reële kans dat retailers in de toekomst geen digitale folder meer zullen uitbrengen. Daarom is het volgens Marc belangrijk om relevant te blijven door middel van de verspreiding en inzichten in data.

Conclusie SWOT-Analyse

Wanneer er gekeken wordt naar de vier facetten van de SWOT-analyse wordt duidelijk dat de fusie aanzienlijke kennis heeft samengebracht met betrekking tot de digitale folder. Niettemin is het van cruciaal belang dat Publitas in eerste instantie de processen, tools en architectuur consolideert. Dit legt de basis voor het benutten van de volgende grote kansen.

De toekomstige kansen liggen voornamelijk op het gebied van data. Voor Publitas is het cruciaal om niet uitsluitend afhankelijk te zijn van het publiceren van digitale folders. Het is essentieel dat retailers, naast het publiceren van deze folders, ook de aanvullende waarde van Publitas erkennen. Hierbij valt te denken aan het faciliteren van de verspreiding van digitale folders naar diverse folderplatformen, het verstrekken van diepgaande data-inzichten op zowel het eigen platform als op de folderplatformen en het ondersteunen van retailers bij de inhoud van de digitale folder met behulp van machine learning.

Bijlage 5 – Interview Toby Nijboer

Interviewer: Nick Reintjes

Geïnterviewde: Toby Nijboer

Functie Geïnterviewde: Back-end developer WePublish

Nick: Hoi Toby, fijn dat ik je vandaag een aantal vragen kan stellen voor mijn afstudeerscriptie. Om je kort wat inzicht te geven in mijn scriptie: ik doe onderzoek naar hoe WP en Publitas op een goede manier de twee IT-architecturen kunnen consolideren. Het uiteindelijke doel is om een ontwerp/advies uit te brengen. Ik heb de architecturen al gemodelleerd binnen Archimate en wil deze nu ook kort met jou doorlopen, met name vanuit het perspectief van WePublish. Waar ik nu naar op zoek ben in dit interview is wat meer inzicht in hoe de schaalbaarheid, datakwaliteit en prestaties worden gewaarborgd.

Zou je om mee te starten je kort kunnen voorstellen en beschrijven wat je functie is bij WePublish/Publitas.

Toby: Nou, ik ben Toby Nijboer, backend developer bij WePublish. Ik ben verantwoordelijk voor alles met betrekking tot het FMP en de viewers aan de WePublish zijde. Daarnaast heb ik het afgelopen jaar me ook beziggehouden met de IT-architectuur vanwege onderbezetting.

Eerste onderwerp datakwaliteit:

Nick: Hoe heeft de organisatie haar IT-architectuur opgebouwd om te zorgen voor optimale datakwaliteit?

Toby: In 2022 ontdekten we dat de IT-architectuur niet schaalbaar was en niet meer naar behoren functioneerde. Toen hebben we de organisatie Figurs ingeschakeld en zij hebben de huidige omgeving opgezet binnen Microsoft Azure. Dit verliep het eerste jaar goed, maar nu stuiten we steeds vaker op problemen in de Architectuurlaag, voornamelijk vanwege een gebrek aan kennis. We proberen het schip boven water te houden, maar daar is voor nu ook alles mee gezegd.

Gelukkig is Gabor Zelei nu begonnen en hebben we wat meer kennis in huis. Echter moet hij nog wel ingewerkt worden.

Nick: Dus momenteel is er een gebrek aan kennis en staat het eigenlijk stil?

Toby: Ja, zo kan je het stellen ja.

Nick: Duidelijk, misschien dat sommige vragen dan lastig te beantwoorden zijn, maar dan kunnen we hier altijd nog op terug komen.

De volgende is: welke maatregelen en processen heeft de organisatie geïmplementeerd om de integriteit en nauwkeurigheid van de data te waarborgen?

Toby: Eigenlijk geen. Pas wanneer de data-analisten problemen tegenkomen binnen PowerBI of in de databases, wordt er gekeken waar het misgaat. Op dat moment kan er echter alleen worden teruggevallen op de ruwe data.

Nick: Duidelijk, verre van een ideale situatie dus. En kun je uitleggen hoe de datakwaliteit binnen de organisatie wordt gemeten en beheerd?

Toby: Net als hierboven gebeurt dit eigenlijk niet binnen de architectuur. Dit wordt pas binnen Power BI gedaan. Wanneer er iets misgaat, wordt pas aan de bel getrokken en niet daarvoor.

Nick: Duidelijk.

Tweede onderwerp data Schaalbaarheid:

Nick: Hoe schaalte de organisatie haar IT-architectuur om te voldoen aan de groeiende behoefte aan gegevensverwerking en -opslag?

Toby: Minimaal, de fact table blijft maar groeien totdat er zelf aan de handrem wordt getrokken. Verder wordt in de pipelines veel gewerkt met statische waarden en niet met dynamische waarden. Dit blijft dus groeien en is niet schaalbaar.

Eigenlijk wordt alles maar aan elkaar geknoopt in de hoop dat het goed gaat en wanneer er iets misgaat, wordt er pas naar gekeken.

Nick: Dat klinkt niet ideaal, zijn hier al veel problemen door ondervonden?

Toby: Ja hier zijn al veel issues naar voren gekomen, echter door goede verwachting management van de Business developers en sales komt het vaak goed.

Nick: Welke uitdagingen heeft de organisatie ondervonden bij het opzetten van een schaalbare IT-architectuur en hoe zijn deze aangepakt?

Toby: Voorheen duurde de PowerBI-refresh ongeveer 2 uur. Na de nieuwe opzet door Figurs is dit teruggebracht naar 10-15 minuten. Maar wat we nu zien in de architectuur is dat het op veel plekken weer misgaat. Dit lijkt mij voornamelijk te wijten aan ontbrekende kennis en we hopen dat dit verbetert nu Gabor is aangenomen.

Nick: Op welke manier worden nieuwe gegevensbronnen geïntegreerd in de bestaande IT-architectuur en hoe wordt de consistentie gewaarborgd?

Toby: In de afgelopen jaren zijn alleen nieuwe externe platformen gekoppeld. Dit gebeurt via zelfgebouwde API's. Echter zijn we met deze bronnen altijd sterk afhankelijk van welke data zij aanbieden en op welke manier. Het is dus altijd een hele klus om dit goed weg te zetten in bruikbare vorm in de databases.

Derde onderwerp data beveiliging

Nick: Hoe wordt de beveiliging van gegevens binnen de IT-architectuur aangepakt en welke maatregelen zijn genomen om gegevens lekken te voorkomen?

Toby: Voor data beveiliging hebben we een aantal zaken ingeregeld. Ten eerste slaan we geen persoonsgegevens op. Verder gebruiken we voor de database login een gebruikersnaam en wachtwoord met specifiek toegekende rechten per gebruiker. Bovendien moet je IP-gewhitelist zijn om in te kunnen loggen. Voor Azure staat 2FA aan, naast de gebruikelijke gebruikersnaam/wachtwoord. En tot slot is er Row Level Security ingesteld om de retailers alleen de correcte data te laten zien die zij mogen zien.

Nick: Duidelijk, hier is dus wel goed over nagedacht.

Toby: Ja, gelukkig wel.

Nick: Hoe wordt monitoring momenteel geïmplementeerd binnen de IT-architectuur

Toby: Grotendeels handmatig, dagelijks wordt gecontroleerd of de pipelines gedraaid hebben in Azure data factory of dat er foutmeldingen zijn. Daarnaast worden de meldingen centra van tools zoals BigQuery gebruikt.

Nick: Duidelijk, verhaal. Als ik het goed samenvat werkt het systeem echter is er een aanzienlijke ruimte voor verbeteringen. Maar daarvoor is nu een nieuwe Data engineer aangenomen.

Toby: Ja, gelukkig wel.

Vierde onderwerp: Archimate tekening

Nick: Tot slot heb ik hier de tekening die ik alvast had opgestuurd, samen met de documentatie die ik graag wil laten controleren. Mijn vraag is nu of deze tekening compleet is en zo niet, waar nog aanpassingen nodig zijn.

Toby: Ik heb een paar kleine opmerkingen. Je hebt de aggregatie stap als service binnen Cosmos. Dit moet een functie zijn. Verder klopt de linkerkant. De MySQL, ETL en MsSQL kloppen ook. Echter, mis ik bij de API's nog Reclamefolder en AlleFolders. Deze moeten hier nog aan toegevoegd worden. Verder geloof ik de PowerBI omgeving wel.

Nick: Dus hoe het proces verloopt, dus de flow is correct weergegeven?

Toby: Ja dit klopt, verder klopt de beschrijving ook. mocht je dit nog in detail moeten beschrijven kunnen we hier nog verder naar kijken. Maar hoog over als Architectuurlaag klopt het.

Nick: Fijn en ook meteen bedankt voor je tijd. Ik ga het interview uitwerken en de smalltalk ertussen uit halen. Daarna zal ik het met je delen. Als je nog iets opvalt tijdens het lezen, hoor ik het graag!

Toby: Geen probleem en succes verder.

Bijlage 6 – Interview Umer Khan

Interviewer: Nick Reintjes

Geïnterviewde: Paul Raya

Functie Geïnterviewde: Data lead Publitas

Nick: Hi Umer, great to have the opportunity to ask you a few questions today for my graduation thesis. To give you a brief insight into my thesis: I am researching how WP and Publitas can effectively consolidate their two data architectures. The ultimate goal is to provide a design/recommendation. I have already modelled the architectures within Archimate and would like to briefly go through them with you, particularly from the perspective of Publitas. What I am currently seeking in this interview is a better understanding of how scalability, data quality, and performance are ensured.

To kick things off, could you please briefly introduce yourself and describe your role at Publitas?

Umer: Hi, I am Umer Khan. I work for Publitas as a DevOps specialist, fully remote. I am the one responsible for the IT-architecture, security, and reliability of all the tooling Publitas uses.

Eerste onderwerp datakwaliteit:

Nick: How has the organization constructed its data architecture to ensure optimal data quality?

Umer: We collect data through Snowplow data tracker. To ensure data quality, we retrieve and initially store the raw data. Subsequently, the data is enriched. During this process, the data is also checked to ensure it aligns with the defined parameters for which it is being collected, using a schema registry. Once the data is validated, it is forwarded; if not, it is directed to the bad stream.

Nick: So, if I understand correctly, you collect the raw data as an archive and a checked version as the data that will be used further in the process?

Umer: Yes, and we enrich the 'checked' version with GeoIP, for example.

Nick: Oke, that is clear! My next question would have been how you ensure the quality of the data. But if I am correct, you do this through the schema registry, correct?

Umer: Yes.

Tweede onderwerp data Schaalbaarheid:

Nick: Oke, so let us move on to scalability. How does the organization scale its data architecture to meet the growing demand for data processing and storage?

Umer: Yeah, for example, not just with this particular service, but with every service we create, we always prioritize high reliability, easy scalability, and security. In this case, you can already observe how we facilitate and ensure scalability, as demonstrated by what we are doing here. This entire system collects approximately one billion events per month and is capable of handling three to four times that amount. We can easily scale up individual components to accommodate this volume.

The key to ensuring scalability is our approach. As mentioned earlier, the collector's primary function is to collect raw events and push them into a data stream. It does not perform any additional tasks. By doing so, we effectively offload the computational heavy lifting from the collector, contributing to the overall reliability and scalability of the system.

Nick: So how do you know you need to upscale the architecture or does this happen automatically? And how is this monitored?

Umer: Everything is monitored, and when a system exceeds a specific value, we receive a notification via email and Slack. When the systems reach capacity, we can immediately add new

resources. The thresholds are set in a way that allows us enough time to scale after receiving a notification.

Nick: Sounds clear, and what challenges has the organization faced in establishing a scalable data architecture, and how have these challenges been addressed?

Umer: From the outset, this system was intentionally designed with scalability in mind. The architecture was constructed with the expectation that it should be capable of handling three, four, or even ten times the anticipated load. To illustrate, we conducted benchmark testing with a billion events per month before proceeding to production. This involved thorough testing and analysis to identify the necessary components and determine the types of instances required to effectively process the data.

It was only after this comprehensive benchmarking process that we moved forward. Fast forward to last November during Black Friday, a peak traffic period, and we witnessed the system effortlessly managing four times the anticipated load. This not only validated the foresight incorporated into the initial design but also demonstrated the robustness of our architecture under real-world, high-demand scenarios.

Nick: In what way are new data sources integrated into the existing data architecture, and how is consistency ensured?

Umer: We have not introduced any new data sources; it is solely our own viewer.

Derde onderwerp data beveiliging

Nick: How is data security addressed within the data architecture, and what measures have been taken to prevent data leaks?

Umer: So, for the collection part, everything is posted privately. What I mean by that is, all these services and components are within a private Virtual Private Cloud (VPC), ensuring they are not exposed externally. The only publicly accessible element is the load balancing. This design secures the system from external intrusions, managed at the architectural level. While there are additional options that we are not currently implementing, such as certain security measures, the existing setup only allows for public access to the specified endpoint. For example, there are no other publicly available services.

We have implemented a concept we refer to as "reducing the blast radius." In this sense, even if one element is compromised, an attacker cannot extend beyond that point. For instance, although the collector endpoint is exposed to the world to enable data posting, we are vigilant about monitoring Snowplow latest updates and security patches. There is a systematic process in place to ensure regular updates for all components, particularly the collector, which is somewhat exposed to the public. The rest of the infrastructure is internal, and the data eventually lands in a secure S3 bucket.

The database incorporates a functionality like GitHub secrets, providing its version of secure credential management. All pertinent credentials are stored as secrets, ensuring that none of the notebooks or pipeline jobs contain hard-coded credentials. This practice enhances security and prevents inadvertent exposure of sensitive information.

Additionally, we employ IP whitelisting to enhance access control. Specifically, we whitelist specific IPs, enabling access to our production databases exclusively for Databricks instances associated with those approved IPs. This adds an extra layer of security by restricting access to trusted sources.

Nick: How is monitoring currently implemented within the data architecture?

Umer: On the production database, detailed logs are maintained to track user activities. This logging mechanism allows us to monitor and trace who is performing what actions. In the event of any suspicious activities, we can investigate and address concerns, providing an additional layer of security and accountability.

Nick: Sounds clear!

Vierde onderwerp: Archimate tekening

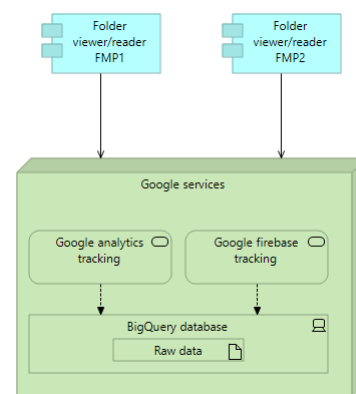
Nick: And finally, I would like to briefly discuss the ArchiMate drawings I sent yesterday. Do you have any feedback on them, or does everything seem accurate?

Umer: Your models look good and accurate. I am not familiar with the connections in ArchiMate, but the Snowplow and Databricks drawings seem correct to me.

Bijlage 7 – IT-architectuur WePublish

Folder viewer en tracking

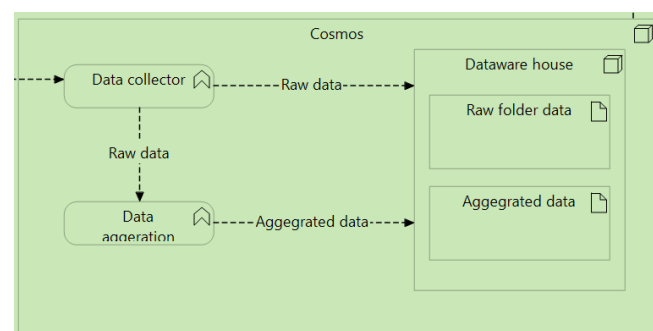
Binnen de applicatie laag zijn de twee WePublish viewers opgenomen: de FMP1 Viewer en de FMP2 Viewer. Van beide viewers wordt de folder data getrackt. De web data wordt getrackt via Google Analytics 4 en de app data via Google Firebase. Deze data wordt vervolgens opgeslagen in BigQuery.



Figuur 25 WP IT-architectuur folder viewer en tracking

Cosmos database

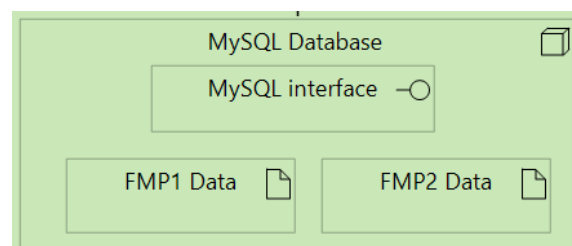
Vervolgens maakt WePublish gebruik van een Cosmos-database die zich bevindt in de Azure Data Factory-omgeving van WePublish. Cosmos haalt de ruwe data op vanuit BigQuery en slaat deze op. Daarnaast worden dezelfde gegevens al geaggregeerd door een aggregatiefunctie, waarna deze geaggregeerde data opnieuw wordt opgeslagen binnen de Cosmos-database. Kort samengevat worden zowel de ruwe data als de geaggregeerde data hier opgeslagen.



Figuur 26 WePublish IT-architectuur Cosmos

MySQL generieke folder data

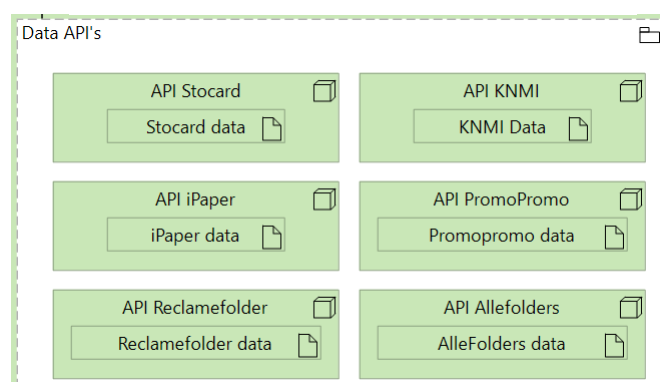
WePublish heeft binnen de Azure Data Factory-omgeving een MySQL-database draaien. Binnen deze database worden alle generieke foldergegevens opgeslagen. Met generieke folderdata wordt bedoeld op de publicaties Pdf's, de URL's, producten, enzovoort.



Figuur 27 WePublish IT-architectuur MySQL

Data API's

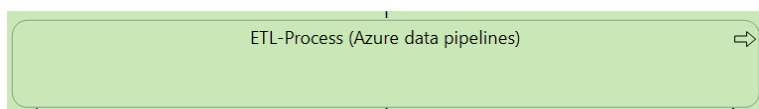
Binnen de Azure Data Factory zijn verschillende API's gebouwd die externe data ophalen van folderplatformen en weerdata van het KNMI. Een belangrijk aspect van de folderplatformen is dat WePublish sterk afhankelijk is van hen. WePublish is namelijk afhankelijk van de data die deze platformen aanbieden en de manier waarop zij dit doen.



Figuur 28 Data API's

Azure data pipelines

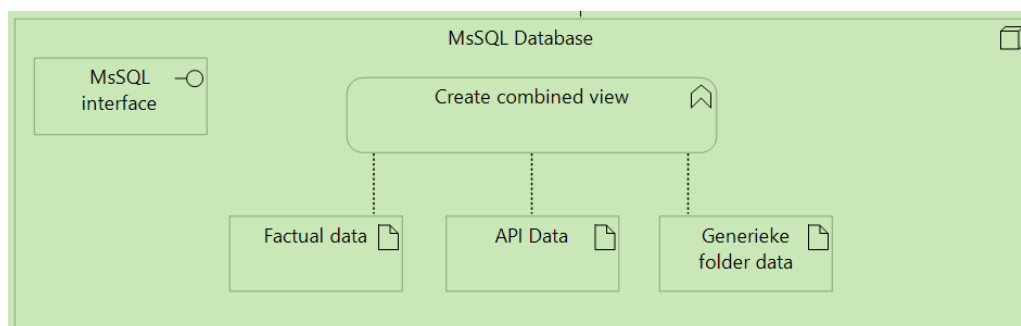
Binnen de Azure Data Factory-omgeving is een ETL-proces opgebouwd om data op te halen van de bronnen Cosmos, MySQL en de data API's. Deze data wordt geaggregeerd en doorgestuurd naar de MsSQL-database.



Figuur 29 WePublish IT-architectuur Azure data pipelines

MsSQL

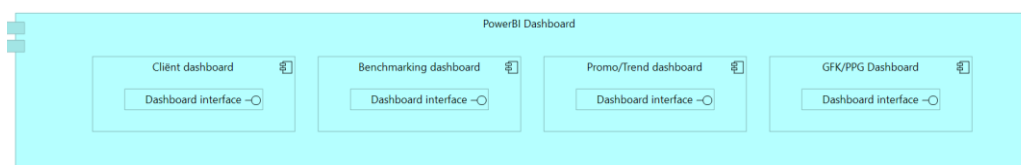
Binnen de Azure Data Factory-omgeving bevindt zich ook een MsSQL-database. In deze database wordt alle data verzameld via het ETL-proces. Van deze gegevens wordt vervolgens een view gemaakt waarin alle data gecombineerd wordt. De feitelijke data wordt gekoppeld aan de generieke folderdata en daarnaast wordt ook de externe API-data hierin verwerkt.



Figuur 30 WePublish IT-architectuur MsSQL database

Dashboarding

WePublish beschikt over een PowerBI-dashboard. Vanuit deze omgeving wordt de vereiste data opgehaald vanuit de MS SQL-server. Hier wordt de geaggregeerde data vanuit de view opgeroepen. Met deze data worden de vier verschillende dashboards gevisualiseerd.



Figuur 31 WePublish IT-architectuur dashboarding

Bijlage 8 – Data-mapping WePublish & Publitas

WePublish data tracking

WePublish maakt gebruik van Google Analytics 4 voor het volgen van de data binnen de WePublish folder viewer. GA4 traceert hierbij de feitelijke gegevens.

Om deze data goed in kaart te brengen, is via een GA4 debugger de data vastgelegd en genoteerd. Hiervoor is de folder geopend, een event getriggerd en deze ruwe data is hieronder opgenomen. Per relevant event is een tabel opgesteld met 3 kolommen. Kolom 1 geeft de naam weer van het event, kolom 2 de trigger (wanneer het event wordt afgevuurd) en kolom 3 bevat de relevante parameters die worden bijgehouden en worden verstuurd naar de database.

Tabel 6 Data tracking WePublish (opgedeeld per event)

Event	unique_view	
Triggers	Een Unique view wordt getriggerd wanneer een bezoeker de digitale folder opent. Dit kan via het platform zelf zijn, maar ook via een directe link naar bijvoorbeeld pagina 13.	
Parameters	page_id	Integer
	page_number	Integer
	publication_id	String
	retailer_id	String
	variant_id	String
	version	Integer
	version_id	String
	view	String

Event	page_visit	
Triggers	Een Page_visit wordt getriggerd wanneer je een folder pagina bezoekt. Dit event wordt alleen verstuurd wanneer je doorklikt naar een volgende pagina.	
Parameters	page_id	Integer
	page_number	Integer
	publication_id	String
	retailer_id	String
	variant_id	String
	version	Integer
	version_id	String

	view	String
--	------	--------

Event	page_leave	
Triggers	Een page_leave wordt getriggert zodra je een pagina verlaat. Een belangrijk detail is het feit dat er page_view_time wordt meegestuurd, op basis van deze metric wordt de pagina leestijd berekent.	
Parameters	page_id	Integer
	page_number	Integer
	page_view_time	Integer
	publication_id	String
	retailer_id	String
	variant_id	String
	version	Integer
	version_id	String
	view	String

Event	web_click	
Triggers	Een web_click wordt getriggert wanneer een klikbaar vak in de digitale folder wordt aangeklikt die doorverwijst naar een website	
Parameters	interaction_id	String
	page_id	Integer
	page_number	Integer
	retailer_id	String
	publication_id	String
	variant_id	String
	version	Integer
	version_id	String
	view	String

Event	image_web_click	
Triggers	Een Image_web_click wordt getriggerd wanneer een afbeelding met link wordt aangeklikt.	
Parameters	interaction_id	String
	page_id	Integer
	page_number	Integer
	retailer_id	String
	publication_id	String
	variant_id	String
	version	Integer
	version_id	String
	view	String

Event	image_page_click	
Triggers	Een image_page_click wordt getriggerd wanneer een afbeelding met link wordt aangeklikt die doorverwijst naar een pagina in de folder.	
Parameters	interaction_id	String
	page_id	Integer
	page_number	Integer
	retailer_id	String
	publication_id	String
	variant_id	String
	version	Integer
	version_id	String
	view	String

Event	image_mail_click	
Triggers	een image_mail_click wordt getriggerd wanneer een afbeelding met een e-mailadres wordt aangeklikt.	
Parameters	interaction_id	String

	page_id	Integer
	page_number	Integer
	retailer_id	String
	publication_id	String
	variant_id	String
	version	Integer
	version_id	String
	view	String

Event	mail_click	
Triggers	Een mail_click wordt getriggerd wanneer een link met een e-mailadres wordt aangeklikt	
Parameters	interaction_id	String
	page_id	Integer
	page_number	Integer
	retailer_id	String
	publication_id	String
	variant_id	String
	version	Integer
	version_id	String
	view	String

Event	video_click	
Triggers	Een Video_click wordt getriggerd wanneer een video in de folder wordt aangeklikt.	
Parameters	interaction_id	String
	page_id	Integer
	page_number	Integer
	retailer_id	String
	publication_id	String
	variant_id	String

	version	Integer
	version_id	String
	view	String

Event	page_click	
Triggers	Een Page_click wordt getriggerd wanneer een link wordt aangeklikt die doorverwijst naar een pagina binnen de folder.	
Parameters	interaction_id	String
	page_id	Integer
	page_number	Integer
	retailer_id	String
	publication_id	String
	variant_id	String
	version	Integer
	version_id	String
	view	String

Event	note_dialog_click	
Triggers	Dit event wordt getriggerd wanneer een interactie wordt aangeklikt die een dialoog opent met een notitie.	
Parameters	interaction_id	String
	page_id	Integer
	page_number	Integer
	retailer_id	String
	publication_id	String
	variant_id	String
	version	Integer
	version_id	String
	view	String

Event	product_dialog_click	
Triggers	Dit event wordt getriggerd wanneer een interactie wordt aangeklikt die een dialoog opent met een product.	
Parameters	interaction_id	String
	page_id	Integer
	oage_number	Integer
	retailer_id	String
	publication_id	String
	variant_id	String
	version	Integer
	version_id	String
	view	String

Event	video_dialog_click	
Triggers	Dit event wordt getriggerd wanneer een interactie wordt aangeklikt die een dialoog opent met een video.	
Parameters	Interaction_id	String
	page_id	Integer
	page_number	Integer
	retailer_id	String
	publication_id	String
	variant_id	String
	version	Integer
	version_id	String
	view	String

Event	webpage_dialog_click	
Triggers	Dit event wordt getriggerd wanneer een interactie wordt aangeklikt die een dialoog opent met een webpagina.	
Parameters	interaction_id	String

	page_id	Integer
	page_number	Integer
	retailer_id	String
	publication_id	String
	variant_id	String
	version	Integer
	version_id	String
	view	String

Event	share_popup_click	
Triggers	Dit event wordt getriggerd wanneer een interactie wordt aangeklikt die een pop_up opent met share mogelijkheden.	
Parameters	interaction_id	String
	page_id	Integer
	page_number	Integer
	retailer_id	String
	publication_id	String
	variant_id	String
	version	Integer
	version_id	String
	view	String

Event	page_zoom	
Triggers	Page_zoom wordt getriggerd wanneer een bezoeker in de folder of app inzoomt.	
Parameters	interaction_id	String
	page_id	Integer
	Page_number	Integer
	retailer_id	String

publication_id	String
variant_id	String
version	Integer
version_id	String
zoom_level	Integer
view	String

WePublish database

De data die via Google Analytics 4 wordt getrackt, wordt eerst opgeslagen binnen BigQuery. Vervolgens wordt de data doorgestuurd naar Cosmos, waar de ruwe gegevens worden opgehaald en weggeschreven in een MsSQL-database door middel van een Azure Data Factory data-pipeline. In dit hoofdstuk zal ik deze geaggregeerde vorm verder uitwerken. Ook zal ik de view behandelen waarin alle data wordt samengevoegd voor gebruik in het dashboard.

Een belangrijk detail om te weten is dat niet alle parameters worden meegestuurd naar de MsSQL-server. Deze data wordt wel in de ruwe vorm opgeslagen in BigQuery. Toby Nijboer gaf aan dat deze data soms voor consultancydoeleinden wordt gebruikt; echter, moet deze data dus worden opgezocht in BigQuery en niet bijvoorbeeld in PowerBI.

Event: Unique views

De Unique View Events van de FMP1- en FMP2-viewers, evenals het folderplatform Folders.nl, worden opgeslagen in [dbo].[WP_Folderviews_Fact]. Om onderscheid te maken tussen de viewers, is view_id opgenomen.

Figuur 33 WePublish events (opgedeeld per event)

	variant_id	date	event_category	total_events	event_action	view_id	retailer_id
1	608b0efd-0970-4f82-a004-a0b106bab883	2023-03-24	folder_view	410	unique_view	rf_firebase_folderdata_v4	d1ec7ad0-8b33-489d-8184-83384659f319
2	82b5b966-438b-4bf1-a8ba-b92320b938da	2022-09-04	folder_view	7	unique_view	rf_firebase_folderdata_v4	3cd8eb55-1e79-4457-a6eb-3d69155ab0dc
3	c24f39af-8fa2-439a-b4f2-ed72bd5c3064	2023-01-28	folder_view	4	unique_view	rf_firebase_folderdata_v4	29cbc0d5-6894-45c2-ac93-5bc8af4d0d22

Event: Page visits en Page leave

De page_visit en page_leave worden opgeslagen in [dbo].[WP_Paginaviews_Fact]. Hier is ook de voorganger van de page_visit te vinden, dit event heet page_hit. Beide gebeurtenissen worden in PowerBI samengevoegd onder één naam.

	variant_id	date	page_number	total_events	page_view_time	page_id	event_action	view_id	retailer_id
50	b662105e-57c4-4436-8eb2-c2d9b19d46de	2023-10-25	7	5	41	efb00d40-10df-42c7-bae3-af41afbedcfa	page_leave	fni-web	3e773456-74b7-48c5-ae7a-af400186ded4
51	52d4d8da-5625-480d-acc6-78ea1b806e6c	2023-10-25	28	29	NULL	06bb9b83-5fb5-4104-ae19-b3a2930c55ec	page_visit	fni-web	babb1e5e-6dd1-4144-89b3-e9134c13289
52	52df3869-ac24-4fcc-bf21-c791b47c5a24	2023-10-23	17	1	NULL	a21e1ffc-e50e-427f-89d0-7e23c31cc05d	page_visit	ok-app	4c570b01-dd9a-4f94-b5e8-d4bcdcb24ab03

Event: Internal- en external clicks

Alle soorten externe clicks en de interne clicks worden opgeslagen in de tabel [dbo].[WP_Clicks_Fact]. Op basis van de event_action wordt in Power BI gefilterd naar een specifiek type, indien nodig.

variant_id	date	total_events	page_id	view_id	interaction_id	event_action	retailer_id
330	8ea05fc2-4e8e-45f3-9641-4f0934c2581	2023-12-22	1	d67e7520-88f3-49f2-b8ea-0a872f8c7363	fnl-web	423b56e3-d341-4271-892f-3437de9185d	d1ec7ad0-8b33-489d-8184-8331
331	c41d43a6-7bbd-49e4-997b-aa23adaa85d9	2023-12-22	2	2cd0851f-c74e-4e9e-a93f-3e6c9f5d3981	wp-viewer-legacy	2e18e2a7-9d33-4d6d-8345-08bcb118c12c	8079
332	0c1512e3-5f04-41a8-9d72-87f2cd41f415	2023-12-22	1	edfa3b2a-201d-4e0f-b0e-ea3d393c3ec	wp-viewer	392d0edd-db14-40f8-857f-f1f44498012f-1	07690d34-353a-4f30-bbcc-ca96
333	a42ca441-d1a6-4b58-98f-26edd7601028	2023-12-22	2	303be4c2-a0fc-44ab-8509-3753779c4265	wp-viewer	8795fcd3-03e5-4505-8792-5579b71b0497	cad5573c-7246-4ebe-82db-03c

Dimensie tabellen

Naast de feitelijke data zijn ook de dimensietabellen opgenomen in de MsSQL-server. Deze dimensietabellen bevatten beschrijvende informatie van de data, inclusief attributen die de kenmerken ervan beschrijven. Denk hierbij aan attributen zoals de retailer naam, folder naam, URL, de periode waarin een folder live staat, enzovoort. In de onderstaande subhoofdstukken zijn de belangrijkste dimensie tabellen opgenomen.

Retailer dimensie

In de retailer dimensie [dbo].[WP_FMP2_Retailers_Dim] is het retailer_id opgenomen, samen met de bijbehorende retailer_naam.

Figuur 34 WePublish dimensie tabellen

	retailer_id	retailer_name
1	3e773456-74b7-48c6-ae7a-af400186ded4	Action
2	d957d152-df80-4f9c-a9f1-712403a00660	AKO
3	58529d9a-8cea-44c2-b137-8cabe17cdaeb	Albert Heijn

Varianten dimensie

In de dimensie [dbo].[WP_FMP_Variants_Dim] zijn attributen van de digitale folder opgenomen, zoals retailer_name, viewer_url, de periode waarin de folder live mag staan en variant_id, enz.

variant_name	viewer_url	publish_from	publish_to	valid_from	valid_to	variant_id	retailer_id	active_version_id	total_pages	publication_id
1	Vomar week 35 2022	https://publicatie.wepublish.com/vomar-week-35...	2022-08-25	2022-08-29	2022-08-27	2022-09-03	e0e971f1-aea3-4aa1-b95b-d888da5ec45	53b589-b25f-4acd-aa33-1e49a15a7055	30	92316974-1709-409a-9e65-91
2	Woonruimte week 35 2022	https://publicatie.wepublish.com/voonruimte-week...	2022-08-26	2022-09-03	2022-08-28	2022-09-03	e5c94d91-a201-41a8-83f1-918c45ae88f1	d5c4389f-cec0-403d-b34d-288ccba82e62	15	440796cf-ac7e-409b-85a6-9
3	TotaalBED week 35 2022	https://publicatie.wepublish.com/votaalbed-week...	2022-08-27	2022-09-03	2022-08-27	2022-09-03	7ceda5a8-730d-4f33-9120-4f118cc9838	399b111-2272-47ad-aae6-01d77c0e0137	14	f9c36d8b-ecce-49ea-a909-e

Pagina dimensie

In de dimensie [dbo].[WP_FMP2_Pages_Dim] zijn de attributen van de digitale folder opgenomen, waaronder page_id, Page_image_url, page_number, Variant_id, enz.

page_id	page_image_url	page_number	variant_id	retailer_id	page_variant_id
1	pdfD79cf63d-ad6e-4c45-b8f9-540f9537bfe4-1	1	30e5715e-3eab-4db7-9fd4-789746a8b734	c2bb06a4-129d-40a3-817a-c322a100debe	pdfD79cf63d-ad6e-4c45-b8f9-540f9537bfe4-1_30e5715e...
2	ddcdccca-daf6-47ab-901e-7c25c387675	86	8dd0639b-f3f1-4667-b0e5-4ea7cdec8ab6	c4df0437-54d6-454d-b286-441e929500bb	ddcdccca-daf6-47ab-901e-7c25c387675_8dd0639b-f3f1...
3	e3a12489-9788-42f5-8134-afe13cf1da9c	30	a47090f6-dee8-42ed-a996-74dd1d073712	c4df0437-54d6-454d-b286-441e929500bb	e3a12489-9788-42f5-8134-afe13cf1da9c_a47090f6-dee8...

Interactie dimensie

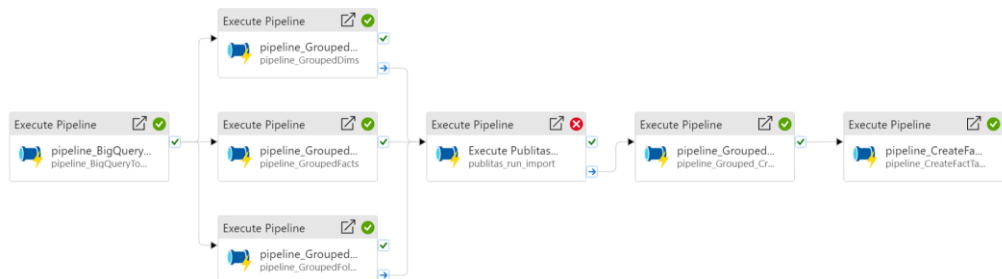
In de tabel [dbo].[WP_FMP2_Interactions_Dim] zijn de interactie-attributen van de folder opgenomen, waaronder interactie_id, page_id, Interaction_type, url en interaction_image_url.

interaction_id	page_id	interaction_type	url	interaction_image_url
1	c1cavagux	hyperlink	http://www.beddenreus.nl/	https://assets.wepublish.com/variant-versions/da...
2	kl4r2p2ow	hyperlink	https://www.beddenreus.nl/boxsprings?SortBy=BestSold	https://assets.wepublish.com/variant-versions/da...
3	mxk48z9dr	hyperlink	https://www.beddenreus.nl/sofabedden/sofabed-dallas-al...	https://assets.wepublish.com/variant-versions/da...

View: Total fact table

Naast dat via de Azure Data Factory pipelines alle data wordt doorgestuurd naar de MsSQL database, is er ook één overkoepelende pipeline opgebouwd die alle verschillende facetten samenvoegt en dit tot één feitentabel realiseert. In onderstaande afbeelding is de volledige serie aan pipelines te zien:

Figuur 35 View total fact table



In de eerste pipeline wordt de BigQuery-data doorgestuurd naar Cosmos. Vervolgens gaan de 3 verschillende 'Grouped' pipelines aan het werk. De eerste zorgt ervoor dat de dimensietabellen worden gegroepeerd, de tweede zorgt ervoor dat de feiten worden samengevoegd en de derde pipeline behandelt de verschillende folderplatformen waar data beschikbaar is.

Wanneer deze gegroepeerde tabellen zijn opgesteld, wordt in de Publitas-pipeline waar mogelijk data toegevoegd vanuit Databricks. Deze pipeline is nog zeer onstabiel, daarom staat deze op "on completion" en niet op "on success". Dit zorgt ervoor dat bij een fout ook de volgende pipeline wordt gestart.

De pipeline 'Grouped_combined' groepeerde alle eerdergenoemde pipeline-uitkomsten, behalve BigQuery naar Cosmos. Tot slot wordt in de pipeline 'CreatefactTable' de view opgesteld die alle samengevoegde data bevat. In onderstaande afbeelding is een voorbeeld te zien van hoe de data eruitziet in de view van de total fact tabel. Deze tabel wordt vervolgens opgehaald door PowerBI en gevisualiseerd.

Figuur 36 Tabel totalfact

	view_id	date	variant_id	page_id	interaction_id	product_id	product_interaction_id	retailer_id	event_action	total_events	page_view_time
92	if_freelance_folderdata_v3	2021-04-05	aaaf0d79-024f-4327-a103-b0c3b745906	78cb8554-533d-48d9-a3ca-e2b8e1e3a3cb	e91917d8-3da7-41c4-bc3e-fb433ce228d8	NULL	e91917d8-3da7-41c4-bc3e-fb433ce228d8	877	click	8	NULL
93	172013858	2021-04-18	e5c5e727-3c38-461a-b97c-a3b385bacc02	aa5a4897-da1b-4567-927c-57ac99a917e	36339a30-05cd-4153-a531-2acc99a917e	NULL	36339a30-05cd-4153-a531-2acc99a917e	7204	click	10	NULL
94	210705942	2021-04-18	43bbc3b8-a950-4b4a-9039-3f1d2330435e	9b6e594-c3f2-4b74-a400-1844b84f391	da93db92-ca2e-412d-90c4-2cc055d8af5e	NULL	da93db92-ca2e-412d-90c4-2cc055d8af5e	907	click	5	NULL

Snowplow data tracking

Zoals eerder vermeld, maakt Publitas gebruik van een Snowplow data tracker. Deze data tracker draait binnen de Publitas folder viewer en traceert de feitelijke gegevens van de bezoeker.

Om deze data goed in kaart te brengen, is via een Snowplow debugger de data vastgelegd en genoteerd. Hiervoor is de folder geopend, een event getriggerd en deze ruwe data is hieronder opgenomen. Per relevant event is een tabel opgesteld met 3 kolommen. Kolom 1 geeft de naam weer van het event, kolom 2 de trigger (wanneer het event wordt afgevuurd) en kolom 3 bevat de relevante parameters die worden bijgehouden en worden verstuurd naar de database.

Tabel 7 Snowplow data tracking (opgedeeld per event)

Event	Unique_view	
Triggers	Een Unique view wordt getriggerd wanneer een bezoeker de digitale folder opent. Dit event kan op elke pagina plaatsvinden, afhankelijk van waar de bezoeker de folder binnenkomt.	
Parameters	Unstructured event	
	event	String
	virtualPageUrl	String
	virtualPageTitle	String
	publitasPageParams	Integer
	Context data	
	publicationId	Integer
	publicationOriginalTitle	Integer
	accountId	String
	accountName	String
	groupId	String
	groupTitle	Integer
	visitId	String



Event	virtualPageView																						
Triggers	Een virtual pageview wordt getriggerd wanneer een bezoeker een folderpagina bezoekt. Als een bezoeker een productpagina opent en vervolgens wegkijkt, wordt er in de Publitas Viewer een nieuwe pageview verstuurd.																						
Parameters	Unstructured event <table> <tr> <td>event</td><td>String</td></tr> <tr> <td>virtualPageUrl</td><td>String</td></tr> <tr> <td>virtualPageTitle</td><td>String</td></tr> <tr> <td>publitasPageParams</td><td>Integer</td></tr> </table> Context data <table> <tr> <td>publicationId</td><td>Integer</td></tr> <tr> <td>publicationOriginalTitle</td><td>Integer</td></tr> <tr> <td>accountId</td><td>String</td></tr> <tr> <td>accountName</td><td>String</td></tr> <tr> <td>groupId</td><td>String</td></tr> <tr> <td>groupTitle</td><td>Integer</td></tr> <tr> <td>visitId</td><td>String</td></tr> </table>	event	String	virtualPageUrl	String	virtualPageTitle	String	publitasPageParams	Integer	publicationId	Integer	publicationOriginalTitle	Integer	accountId	String	accountName	String	groupId	String	groupTitle	Integer	visitId	String
event	String																						
virtualPageUrl	String																						
virtualPageTitle	String																						
publitasPageParams	Integer																						
publicationId	Integer																						
publicationOriginalTitle	Integer																						
accountId	String																						
accountName	String																						
groupId	String																						
groupTitle	Integer																						
visitId	String																						



Event	publitasEvent_internalNavigation																										
Triggers	Dit event wordt getriggerd wanneer een bezoeker één van de twee navigatieknoppen aan de rechter- of onderkant van de folder viewer gebruikt. In de afbeelding rechts zijn deze knoppen uitgelicht.																										
Parameters	Unstructured event <table> <tr><td>event</td><td>String</td></tr> <tr><td>publitasEventParams</td><td>String</td></tr> <tr><td>publitasEventCategory</td><td>String</td></tr> <tr><td>publitasEventAction</td><td>String</td></tr> <tr><td>publitasEventLabel</td><td>Integer</td></tr> <tr><td>publitasEventName</td><td>String</td></tr> </table> Context data <table> <tr><td>publicationId</td><td>Integer</td></tr> <tr><td>publicationOriginalTitle</td><td>Integer</td></tr> <tr><td>accountId</td><td>String</td></tr> <tr><td>accountName</td><td>String</td></tr> <tr><td>groupId</td><td>String</td></tr> <tr><td>groupTitle</td><td>Integer</td></tr> <tr><td>visitId</td><td>String</td></tr> </table>	event	String	publitasEventParams	String	publitasEventCategory	String	publitasEventAction	String	publitasEventLabel	Integer	publitasEventName	String	publicationId	Integer	publicationOriginalTitle	Integer	accountId	String	accountName	String	groupId	String	groupTitle	Integer	visitId	String
event	String																										
publitasEventParams	String																										
publitasEventCategory	String																										
publitasEventAction	String																										
publitasEventLabel	Integer																										
publitasEventName	String																										
publicationId	Integer																										
publicationOriginalTitle	Integer																										
accountId	String																										
accountName	String																										
groupId	String																										
groupTitle	Integer																										
visitId	String																										



Event	publitasEvent_userExit																										
Triggers	Dit event wordt getriggert wanneer een bezoeker de folder verlaat. Het event wordt ook getriggert wanneer een bezoeker een product aanklikt en er een nieuw tabblad wordt geopend. Als de bezoeker vervolgens het producttabblad sluit en teruggaat naar de folder, wordt dit ook weer gerekend als een virtualPageview.																										
Parameters	Unstructured event <table> <tr><td>event</td><td>String</td></tr> <tr><td>publitasEventParams</td><td>String</td></tr> <tr><td>publitasEventCategory</td><td>String</td></tr> <tr><td>publitasEventAction</td><td>String</td></tr> <tr><td>publitasEventLabel</td><td>Integer</td></tr> <tr><td>publitasEventName</td><td>String</td></tr> </table> Context data <table> <tr><td>publicationId</td><td>Integer</td></tr> <tr><td>publicationOriginalTitle</td><td>Integer</td></tr> <tr><td>accountId</td><td>String</td></tr> <tr><td>accountName</td><td>String</td></tr> <tr><td>groupId</td><td>String</td></tr> <tr><td>groupTitle</td><td>Integer</td></tr> <tr><td>visitId</td><td>String</td></tr> </table>	event	String	publitasEventParams	String	publitasEventCategory	String	publitasEventAction	String	publitasEventLabel	Integer	publitasEventName	String	publicationId	Integer	publicationOriginalTitle	Integer	accountId	String	accountName	String	groupId	String	groupTitle	Integer	visitId	String
event	String																										
publitasEventParams	String																										
publitasEventCategory	String																										
publitasEventAction	String																										
publitasEventLabel	Integer																										
publitasEventName	String																										
publicationId	Integer																										
publicationOriginalTitle	Integer																										
accountId	String																										
accountName	String																										
groupId	String																										
groupTitle	Integer																										
visitId	String																										

Event	publitasEvent_linkClicks																										
Triggers	Dit event wordt getriggerd wanneer een product vak wordt aangeklikt. Onder dit event vallen verschillende soorten clicks, waarbij het type click wordt meegegeven in publitasEventAction. Hier vallen bijvoorbeeld standaard URL-clicks, pop-up-clicks of e-mail clicks onder																										
Parameters	Unstructured event <table> <tr> <td>event</td><td>String</td></tr> <tr> <td>publitasEventParams</td><td>String</td></tr> <tr> <td>publitasEventCategory</td><td>String</td></tr> <tr> <td>publitasEventAction</td><td>String</td></tr> <tr> <td>publitasEventLabel</td><td>Integer</td></tr> <tr> <td>publitasEventName</td><td>String</td></tr> </table> Context data <table> <tr> <td>publicationId</td><td>Integer</td></tr> <tr> <td>publicationOriginalTitle</td><td>Integer</td></tr> <tr> <td>accountId</td><td>String</td></tr> <tr> <td>accountName</td><td>String</td></tr> <tr> <td>groupId</td><td>String</td></tr> <tr> <td>groupTitle</td><td>Integer</td></tr> <tr> <td>visitId</td><td>String</td></tr> </table>	event	String	publitasEventParams	String	publitasEventCategory	String	publitasEventAction	String	publitasEventLabel	Integer	publitasEventName	String	publicationId	Integer	publicationOriginalTitle	Integer	accountId	String	accountName	String	groupId	String	groupTitle	Integer	visitId	String
event	String																										
publitasEventParams	String																										
publitasEventCategory	String																										
publitasEventAction	String																										
publitasEventLabel	Integer																										
publitasEventName	String																										
publicationId	Integer																										
publicationOriginalTitle	Integer																										
accountId	String																										
accountName	String																										
groupId	String																										
groupTitle	Integer																										
visitId	String																										



Databricks database

In dit hoofdstuk wordt dieper ingegaan op hoe de eventstructuur eruitziet in de Databricks-database. Dit is onderverdeeld in twee delen: de eventdata vanuit Snowplow en de dimensiedata, wat de kenmerken zijn van de feitelijke data.

Event data

In dit hoofdstuk wordt gekeken naar de drie tabellen waarin de feitelijke data wordt opgeslagen. Hier zal per tabel één rij met conceptuele data worden getoond om inzicht te bieden in hoe deze data eruitziet. Verder zal kort worden toegelicht waarvoor de data in de tabel dient. Deze data wordt in het dashboarding verrijkt met dimensiedata om zo meer waarde aan de data te geven.

Unique view

In de 'unique view' tabel wordt op dag niveau het aantal unieke folderweergaven opgeslagen. Hier worden per retailer en publicatie het aantal views getoond, maar ook het aantal gelezen pagina's en hoeveel bezoekers naast de unique view geen event triggeren binnen de folder, dus een 'bounce' zijn.

De meest relevante kolommen hierbij zijn `account_id`, `publicatie_id`, `event_date` en het totale aantal gebeurtenissen. Met deze inzichten kan de waarde al goed worden weergegeven in de dashboards.

app_id	event_date	account_id	account_name	group_id	group_title	collection_id	collection_title	publication_id	publication_title	publication_view_counts	publication_visit_counts	publication_bounce_counts
reader-prod	10-12-2022	9	METRO AG	576	Doel+D&D D1 D' null	null		1564981	0'N'00,0'0'0'0'0'0,0,0'N'00'N'0'0'0,	26875	1977	252

Figuur 37 Databricks unique views

Pageview

In de 'page view' tabel worden op dag niveau de paginaviews opgeslagen per pagina. Dus het aantal paginaviews per pagina per dag. De meest relevante kolommen zijn hier `event_id`, `event_date`, `account_id`, `publication_id` en `pagina_id`.

Deze tabel wordt gebruikt in het dashboard om het paginaverloop te visualiseren, maar ook de click-through rate en leestijd per pagina. Dit zijn de meest relevante pagina-inzichten volgens de senior retail consultant.

event_id	event_date	event_id	account_id	account_name	group_id	group_title	collection_id	collection_title	publication_id	publication_title	page_type	page_title	page_index	page_id
000e403-e92e-41d5-8b2f-54c3c59476	19-12-2022	4b248f37-9b77-4b75-8605-a9011bc054dc	80499	dm drogierie Croatia	84563	dm-croatia	null		1570589	katalog Prosmac 12-12 do 31-12-2022	product	dm-croatia - kati 29-29	8.01E+12	

Figuur 38 Databricks page views

Page events

Naast de 'Page_views' tabel is er ook de 'pagina events' tabel. In deze tabel worden alle soorten events vastgelegd, zoals clicks en interne navigatie. De opdeling hiervan is te vinden in de kolom 'page_event_category'. De tabellen 'account_id', 'publication_id' en 'page_event_params' worden gebruikt om precies te bepalen over welke pagina het gaat.

Samen met de 'page_view' tabel kunnen inzichten zoals CTR% per pagina en clicks per pagina binnen het dashboard inzichtelijk worden gemaakt.

event_date	account_id	account_name	group_id	group_title	collection_id	collection_title	publication_id	publication_title	page_event_type	page_event_params	page_event_category	page_event_action	page_event_label	page_event_name	page_event_count
6-9-2022	49100	catalog	50838	catalog	null		1518189	AucC15M022Folio	publitasEvent_internal/Navigation	["type":"next","url":"https://view.publ internal links	next		54-55	user_navigation	18

Figuur 39 Databricks page events

Dimensies

Naast de feitelijke data heeft Publitas, net als WePublish, dimensietabellen. De dimensietabellen bevatten kenmerken waarmee de feitelijke data wordt verrijkt. Denk hierbij bijvoorbeeld aan tijdsbepaling, retailers, publicatie informatie en productinformatie.

Retailer/Accounts

De eerste dimensietabel is de accountdimensietabel. Hier worden alle beschrijvende data van een retailer bijgehouden, zoals de naam, datum van creatie, locatiegegevens en betalingsgegevens.

Publications

De tweede dimensietabel is de publicatietabel. In deze tabel vind je alle relevante informatie over de publicatie. Voorbeelden van velden zijn titel, creatiedatum en aantal pagina's.

Product

De derde dimensietabel is de producttabel. Hier wordt dieper ingegaan op de producten die in de folder worden getoond, voornamelijk gericht op dynamische producten uit een productfeed. Een productfeed is een feed die de retailer aanlevert, waarvan Publitas vervolgens data haalt op basis van filters en deze toont in de folder. Data zoals titel, positie en prijs worden hierin opgeslagen.

accounts		publications		products	
id	SERIAL	id	SERIAL	id	SERIAL
name	CHARACTER VARYING(255)	title	CHARACTER VARYING(255)	hotspot_id	INTEGER
created_at	TIMESTAMP(6) WITHOUT TIME ZONE	description	TEXT	title	CHARACTER VARYING(150)
updated_at	TIMESTAMP(6) WITHOUT TIME ZONE	created_at	TIMESTAMP(6) WITHOUT TIME ZONE	description	TEXT
contract_publisher_id	INTEGER	updated_at	TIMESTAMP(6) WITHOUT TIME ZONE	position	INTEGER
contract_publisher_instance	CHARACTER VARYING(255)	group_id	INTEGER	created_at	TIMESTAMP(6) WITHOUT TIME ZONE
notes	TEXT	state	CHARACTER VARYING(255)	updated_at	TIMESTAMP(6) WITHOUT TIME ZONE
billing_address	CHARACTER VARYING(255)	slug	CHARACTER VARYING(255)	price	CHARACTER VARYING(255)
billing_zipcode	CHARACTER VARYING(255)	custom_header	TEXT	discounted_price	CHARACTER VARYING(255)
billing_city	CHARACTER VARYING(255)	schedule_online_at	TIMESTAMP(6) WITHOUT TIME ZONE	feed_product_id	BIGINT
billing_country	CHARACTER VARYING(255)	schedule_offline_at	TIMESTAMP(6) WITHOUT TIME ZONE	webshop_url	CHARACTER VARYING(2000)
tax_number	CHARACTER VARYING(255)	used	BOOLEAN	webshop_identifier	CHARACTER VARYING(255)
bank_account	CHARACTER VARYING(255)	online_at	TIMESTAMP(6) WITHOUT TIME ZONE	linked	BOOLEAN
billing_contact_name	CHARACTER VARYING(255)	pages_count	INTEGER	video_url	CHARACTER VARYING(255)
billing_contact_email	CHARACTER VARYING(255)	appeared_at	DATE	feed_product_sha	CHARACTER VARYING
cc_name	CHARACTER VARYING(255)	product_hotspots	BOOLEAN	availability	CHARACTER VARYING
cc_type	CHARACTER VARYING(255)	external_link_hotspots	BOOLEAN	item_group_id	CHARACTER VARYING(50)
cc_expiry_date_month	INTEGER	page_reference_hotspots	BOOLEAN	product_type	CHARACTER VARYING(750)
cc_expiry_date_year	CHARACTER VARYING(255)	video_hotspots	BOOLEAN	brand	CHARACTER VARYING(70)
stripe_id	CHARACTER VARYING(255)	language	CHARACTER VARYING(10)	gtin	CHARACTER VARYING(14)
generated	BOOLEAN	offline_at	TIMESTAMP(6) WITHOUT TIME ZONE	mpn	CHARACTER VARYING(70)
deleted_at	TIMESTAMP(6) WITHOUT TIME ZONE	browser_title	CHARACTER VARYING(255)	custom_label_0	CHARACTER VARYING(100)
current_license_contract_id	INTEGER	first_online_at	TIMESTAMP(6) WITHOUT TIME ZONE	custom_label_1	CHARACTER VARYING(100)
self_service	BOOLEAN	modified_at	TIMESTAMP(6) WITHOUT TIME ZONE	custom_label_2	CHARACTER VARYING(100)
invoice_delivery_method	INTEGER	example	BOOLEAN	custom_label_3	CHARACTER VARYING(100)
automated_invoicing	BOOLEAN	deleted_at	TIMESTAMP(6) WITHOUT TIME ZONE	custom_label_4	CHARACTER VARYING(100)
provisioned	BOOLEAN	deleted_by_id	INTEGER	google_product_category	CHARACTER VARYING
currency	CHARACTER VARYING	cloning	BOOLEAN	library_product_id	BIGINT
moneybird_contact_id	CHARACTER VARYING	show_related_publications	BOOLEAN	file_path	CHARACTER VARYING
disable_automatic_billing_retry	BOOLEAN	layout_type	INTEGER	video_autoplay	BOOLEAN
partner_account	BOOLEAN	secured	BOOLEAN		
verification_requested	TIMESTAMP(6) WITHOUT TIME ZONE	password_digest	CHARACTER VARYING(255)		
custom_header	TEXT	source_document_id	INTEGER		
marked_as_safe	BOOLEAN	auto_related_publications	BOOLEAN		
billing_contact_email_cc	CHARACTER VARYING(255)	really_destroyed_at	TIMESTAMP(6) WITHOUT TIME ZONE		
		disable_seo	BOOLEAN		
		force_single_page_on_mobile	BOOLEAN		
		mobile_vertical_scroll	BOOLEAN		
		override_mobile_layout	BOOLEAN		
		desktop_vertical_scroll	BOOLEAN		

Figuur 40 Dimensie tabellen Databricks

Bijlage 9 – Stakeholder analyse

In de stakeholderanalyse worden de stakeholders van het project geïdentificeerd. Hierbij wordt onderzocht wie belang heeft bij het project en wat de mate van belang is. Het doel is niet alleen het identificeren van de stakeholders, maar ook het inzichtelijk maken van hun communicatiebehoefte en dit nauwkeurig te documenteren, om zo in de toekomst minder weerstand te ondervinden. Verder wordt er gekeken naar de prioritering van de stakeholders.

Identificatie stakeholders

In het hoofdstuk 'Identificatie van stakeholders' wordt vastgelegd wie de interne en externe stakeholders zijn. Deze stakeholders worden geïdentificeerd en er wordt vermeld waarom zij belanghebbenden zijn bij het project en in welke mate. Vervolgens zullen deze stakeholders ook worden geclassificeerd op basis van hun belang voor het project.

Interne stakeholder

Onder interne stakeholders worden alle interne afdelingen benoemd die belang hebben bij het project. Later in de uitwerking van deelvraag 3 worden de wensen in kaart gebracht door middel van interviews met personen uit de betreffende afdeling.

Business consultants

De business consultants zijn stakeholders voor dit project. De reden hiervoor is dat zij actieve relaties onderhouden met de retailers, waarbij de nadruk ligt op het optimaliseren van digitale folders. Om dit te bereiken is data nodig voor prestatie inzichten, trend/promo inzichten en ook A/B-testing. Het is voor de business consultant van belang dat er goed wordt nagedacht over de toekomstige data en de visualisatie daarvan.

Sales medewerkers

De salesmedewerkers van zowel Publitas als WePublish zijn stakeholders voor dit project. De salesmedewerkers gebruiken folderdata om in de eerste plaats retailers binnen te halen, denk hierbij aan benchmarking van de branche, uitkomsten van A/B-testen en historische data. Daarnaast is data ook relevant wanneer er wordt overwogen een retailer up te sellen. Het is dus voor sales van belang dat er goed wordt nagedacht over de toekomstige data en de visualisatie daarvan.

Productieafdeling

De productieafdeling is ook een stakeholder van het project. De productieafdeling maakt gebruik van de inzichten uit de folderdata om te controleren of de folders van belangrijke retailers live staan. Hoewel de productieafdeling een stakeholder is, is deze niet zo cruciaal als de business consultants en de salesafdeling. Zij zullen ook minder inbreng hebben wat betreft de visualisatie.

Management

Management is een stakeholder van het project, zij hebben directe invloed op de koers en de macht om het project te stoppen of te ondersteunen. Als zij de koers van de informatiestrategie aanpassen, zal dit directe invloed hebben op het project.

DevOps

DevOps is een stakeholder, aangezien zij direct betrokken zullen zijn bij de consolidatie van de data en het beheer van verschillende componenten. Bovendien zijn ze verantwoordelijk voor taken zoals monitoring en het waarborgen van de schaalbaarheid van de componenten.

Externe stakeholders

Externe stakeholders omvatten de belanghebbenden die buiten de organisatie Publitas vallen.

Retailers en mediabureaus

Wat betreft externe stakeholders hebben we er twee, namelijk retailers en mediabureaus die de digitale folder beheren voor de retailer. Deze zijn van groot belang in het project, aangezien de uiteindelijke bestemming van de data voor hen is. Echter, omdat we niet met nieuwe inzichten gaan werken maar de data gaan consolideren, wordt er bij het bekijken van de visualisatiemogelijkheden uitgegaan van de interne stakeholders.

Categorisatie Stakeholders

Na de identificatie van de belanghebbenden zijn ze in onderstaande matrix opgedeeld in primaire en secundaire stakeholders.

De primaire stakeholders hebben directe invloed op zowel de aanpak als het resultaat. Dit zijn de business consultants, sales medewerkers en het management. De business consultants en salesmedewerkers zullen het uiteindelijke dashboard actief gaan gebruiken voor sales en consultancy doeleinden. Het management is een primaire stakeholder omdat zij tijd moeten toekennen voor de realisatie.

Als er gekeken wordt naar de secundaire stakeholders, behoren DevOps en de productieafdeling tot de interne stakeholders. Productie zal minder actief gebruikmaken van het dashboard en het alleen ter validatie gebruiken. DevOps krijgt uiteindelijk de verantwoordelijkheid voor het beheer. Verder behoren retailers en mediabureaus tot de externe secundaire stakeholders. Zij zullen data ontvangen en toegang krijgen tot het dashboard. Echter zullen zij niet betrokken worden bij de realisatie en visualisaties.

Stakeholders	Primair Heeft direct invloed op projectaanpak/resultaat	Secundair Heeft indirect invloed op projectaanpak/resultaat
Interne stakeholder Bij het project betrokken vanuit de eigen organisatie	• Business consultants • Sales medewerkers • Management	• Productie afdeling • DevOps
Externe stakeholder Bij het project betrokken vanuit een externe organisatie		• Retailers • Mediabureaus

Figuur 41 Categorisatie Stakeholders

Prioritering, invloed en communicatie stakeholders

Om een goede communicatiestrategie te realiseren, zijn de stakeholders opgenomen in een stakeholdermatrix. In deze matrix wordt gekeken naar de invloed/kracht van de stakeholder en de belangen van de stakeholders. Hieruit komen vier vakken naar voren, namelijk:

- Hoge invloed en hoog belang, dit zijn de sleutelfiguren die nauwlettend in de gaten moeten worden gehouden en waarmee je nauw samenwerkt. In het geval van dit project zijn dit de business consultants en salesmedewerkers.
- Hoge invloed en laag belang, dit zijn de beïnvloeders. In het geval van dit project betreft het management. Het is van belang om deze stakeholder op cruciale momenten te informeren.
- Lage invloed en hoog belang, dit zijn de geïnteresseerden. In dit geval zijn dat de productieafdeling, retailers en mediabureaus. Deze groep dien je te informeren.
- Lage invloed en laag belang, dit zijn de toeschouwers. In dit geval vertegenwoordigd door DevOps. Deze groep vereist over het algemeen minder aandacht.

		Belang van stakeholder			
		Laag	Matig	Hoog	Zeer hoog
Invloed/macht van stakeholder	Zeer hoog	Beïnvloeder <i>Tevreden houden</i> Management		Sleutelfiguur <i>Samenwerken</i> Business consultants Sales medewerkers	
	Hoog				
	Matig	Toeschouwer <i>Monitoren</i> DevOps		Geïnteresseerd <i>Informeren</i> Productie afdeling Retailers Mediabureaus	
	Laag				

Figuur 42 Prioritering stakeholders

Bijlage 10 – Interview Ibe Hundling

Interviewer: Nick Reintjes

Geïnterviewde: Ibe Hundling

Functie Geïnterviewde: Senior retail consultant

Nick: Hoi Ibe, Fijn dat je even tijd hebt voor een aantal vragen voor mijn afstudeerscriptie. Om je kort wat inzicht te geven in mijn onderzoek: ik doe onderzoek naar hoe WePublish en Publitas de twee data-architecturen kunnen consolideren. Hierbij hoort ook de data die wordt bijgehouden vanuit de folder. Ik heb zowel de getrackte data van WePublish als die van de Publitas viewer in kaart gebracht.

Het doel van dit interview is om te achterhalen hoe je tegen de huidige getrackte data en visualisaties aankijkt en wat er goed en minder goed gaat. Deze resultaten worden vervolgens gebruikt om te analyseren hoe het huidige systeem verschilt van het gewenste systeem. Dit wordt gedaan door middel van een GAP-analyse, waarvan dit interview een van de inputbronnen is.

Folder data tracking & Dashboarding

Nick: Ben je over het algemeen tevreden met de mogelijkheden die het dashboard biedt?

Ibe: Ja en nee. Ik ben tevreden met de mogelijkheden die het WePublish-dashboard biedt, echter, de continuïteit laat soms te wensen over en missen we diepgaande inzichten in de Publitas-data. Aan de andere kant ben ik minder te spreken over het Publitas-dashboard, waar de inzichten te basaal zijn en het dashboard traag is.

Wat ik graag zou willen zien, zijn inzichten die meer gericht zijn op productniveau en voorspellende analyses.

Nick: Ja dat begrijp ik. Maar de gewenste inzichten richten zich echter niet direct op de feitelijke data die wordt gemeten vanuit de folderviewer. Als we specifiek naar de getrackte folder data kijken, met name naar de events Folder_view, Page_views, gegroepeerde Clicks, pagina leestijd en de bounce rate, vraag ik me af of er nog andere events zijn die je hier mist?

Ibe: Nee, volgens mij is dit compleet.

Nick: Dat is mooi, echter zitten we wel met een verschil tussen WePublish en Publitas. Ze meten Page_views namelijk anders: als een bezoeker bij Publitas een product aanklikt, een pop-up opent en deze vervolgens sluit, wordt er een nieuwe Page_views getrackt. Bij WePublish gebeurt dit niet. Wat zou jouw voorkeur zijn en waarom?

Ibe: De manier van WePublish, de bezoeker ziet inderdaad 2 keer de pagina, maar dit zal een raar verloop vertonen bij goed geklikte pagina's. Het doel van de Page_views is het verloop van de folder in kaart brengen en dit bereik je niet door het aantal Page_views te verhogen elke keer dat ze worden weggeschoten bij het sluiten van een pop-up.

Nick: Duidelijk. Dan heb ik nog een vraag. In beide viewers hebben we verschillende soorten clicks, denk hier aan dynamische click, pop-up click, standaard click etc. Hoe zie je dit het liefst gevisualiseerd binnen het dashboard? Zou je de voorkeur geven aan een gegroepeerde weergave onder één algemene naam, bijvoorbeeld 'Clicks', met de mogelijkheid om vervolgens te filteren op de achtergrond van het dashboard? Of zou je het liever opgedeeld zien binnen het dashboard?

Ibe: Voor de duidelijkheid van het dashboard zie ik dit liever onder één naam. Het opsplitsen kan verwarrend zijn en tot veel vragen leiden, dat willen we zo veel mogelijk vermijden.

Nick: Top, dat is duidelijk. Verder heb ik geen vragen betreft de tracking. Nu wil ik graag kijken naar het dashboard, ben je over het algemeen tevreden over de huidige dashboarding?

Ibe: Grotendeels wel

Nick: Oke en wat mis je dan precies?

Ibe: Ik ben grotendeels tevreden met het dashboard, maar de opmaak is verouderd omdat het nog in het WePublish-jasje zit. Dit moet overgezet worden naar de blauwe themakleuren van Publitas, aangezien we ook in de toekomst onder die naam verdergaan.

Verder moet de data van Publitas worden toegevoegd op een manier die goed samenwerkt met de data die al in het dashboard staat, dat is nu niet het geval.

Wat ik verder nog kwijt wil, vind ik dat er te weinig kennisdeling is als je kijkt naar data, dashboarding en consultancy. De organisaties liggen nog te ver van elkaar af. CSM weet bijvoorbeeld niet wat allemaal mogelijk is met het dashboard van WePublish.

Nick: Duidelijk en qua visualisaties zal ik onder andere naar de literatuur gaan kijken, maar wat voor visualisaties gaat jouw voorkeur naartoe?

Ibe: Staaf of Lijndiagrammen als het gaat over een verloop van meerdere dagen/weken. Verder vind ik buttons erg duidelijk. Waar ik persoonlijk absoluut geen fan van ben zijn cirkeldiagrammen.

Nick: Dan zal ik in ieder geval geen cirkeldiagrammen aanbevelen.

Dus als ik het mag samenvatten: Je bent over het algemeen tevreden met het dashboard zoals het momenteel bij WePublish staat, vooral als het gaat om de data-inzichten die verband houden met de Folderdata. Echter moet hier Publitas data aan worden toegevoegd en dient de huisstijl aangepast te worden naar een ontwerp gebaseerd op het thema van Publitas.

Daarnaast wil je dat er op dezelfde manier als bij WePublish wordt omgegaan met de page_views en dat de clicks in het dashboard worden gegroepeerd.

Bijlage 11 – Interview Ramon Eerdman

Interviewer: Nick Reintjes

Geïnterviewde: Ramon Eerdman

Functie Geïnterviewde: Sales Publitas, WePublish en Folders.nl

Nick: Hoi Ramon, Fijn dat je even tijd hebt voor een aantal vragen voor mijn afstudeerscriptie. Om je kort wat inzicht te geven in mijn onderzoek: ik doe onderzoek naar hoe WePublish en Publitas de twee data-architecturen kunnen consolideren. Hierbij hoort ook de data die wordt bijgehouden vanuit de folder. Ik heb zowel de getrackte data van WePublish als die van de Publitas viewer in kaart gebracht.

Het doel van dit interview is om te achterhalen hoe je tegen de huidige getrackte data en visualisaties aankijkt en wat er goed en minder goed gaat. Deze resultaten worden vervolgens gebruikt om te analyseren hoe het huidige systeem verschilt van het gewenste systeem. Dit wordt gedaan door middel van een GAP-analyse, waarvan dit interview een van de inputbronnen is.

Zou je om mee te starten je kort kunnen voorstellen en beschrijven wat je functie is bij WePublish/Publitas.

Ramon: Hey, ja, geen probleem. Mijn naam is Ramon Eerdman en ik werk bij de Sales-afdeling van Publitas/WePublish. Ik richt me voornamelijk op de markt van de Benelux. Ik kom in contact met veel klanten die op zoek zijn naar een viewer voor hun eigen platform, maar ik ben ook verantwoordelijk voor de folderplatformen Folders.nl en Promobutler. Verder weet ik veel van productie af omdat ik daar eerst hebt gewerkt en ben doorgesloegen naar de sales afdeling.

Folder data tracking & Dashboarding

Nick: Duidelijk verhaal. Ik heb een aantal vragen voor je die betrekking hebben op het tracken van folderdata en het gebruik van dashboards. Mijn eerste algemene vraag is of je tevreden bent met het huidige dashboard.

Ramon: Ja, eigenlijk wel. Ik maak voornamelijk gebruik van het WePublish-dashboard. De reden hiervoor is dat ik vaak data gebruik om inzicht te krijgen in het bereik en het aantal clicks. Voor meer gedetailleerde inzichten, zoals het paginaverloop en de best presterende pagina's, haak het data-team aan.

Nick: Top en ben je naast deze basisinzichten ook nog opzoek naar verdere inzichten of is dit al voldoende om je taken goed uit te voeren?

Ramon: Ja, dit is in principe voldoende. Waar ik wel tegenaan loop, is dat de opmaak verouderd is. De screenshots die ik gebruik, zijn vaak in strijd met het thema van de slides die ik meeneem naar een vergadering.

Nick: Fijn dat de data in ieder geval voldoende is. Met de consolidatie wordt ook meteen de opmaak van de dashboards meegenomen. Er wordt nu dus ook nagedacht over de vormgeving.

Ramon: Wat bedoel je precies met de consolidatie?

Nick: De samenvoeging van de twee data architecturen van WePublish en Publitas. Maar als toevoeging hiervan wordt ook gelijk nagedacht over toekomstige dashboarding.

Ramon: Ah oké, duidelijk.

Nick: Dan heb ik nog een vraag. In beide viewers hebben we verschillende soorten clicks, denk hier aan dynamische click, pop-up click, standaard click etc. Hoe zie je dit het liefst gevisualiseerd binnen het dashboard? Zou je de voorkeur geven aan een gegroepeerde weergave onder één algemene

naam, bijvoorbeeld 'Clicks', met de mogelijkheid om vervolgens te filteren op de achtergrond van het dashboard? Of zou je het liever opgedeeld zien binnen het dashboard?

Ramon: Dit maakt voor mij niet veel uit, vanuit sales haken we team data aan voor deze inzichten. Hier kan je zelf dus een oplossing voor kiezen.

Nick: Dit is mooi, we hebben van Ibe wel een sterke voorkeur, dus gaan we uit van zijn mening.

Verder heb ik nog een aantal vragen over de dashboarding. Ben je over het algemeen tevreden met de dashboards?

Ramon: Ja, maar zoals ik net aangaf, zie ik graag een ander thema. Verder zou ik in de toekomst ook graag meerdere platformen in het dashboard willen zien. Denk bijvoorbeeld aan Bonial, iPaper of Folderz. Hoe meer data wij kunnen aanbieden, hoe waardevoller we kunnen zijn voor de retailer.

Nick: Dus als ik het goed begrijp vind je de huidige dashboarding al goed. Echter zou je in de toekomst meerdere folderplatformen willen koppelen en ook gelijk het thema willen aanpassen.

Ramon: Ja precies.

Nick: Duidelijk en qua visualisaties zal ik onder andere naar de literatuur gaan kijken, maar wat voor visualisaties gaat jouw voorkeur naartoe?

Ramon: Duidelijk diagrammen die het verloop aantonen. Qua grafieken o.i.d. heb ik geen sterke voorkeur. Voor mij geldt wel hoe overzichtelijker hoe beter.

Nick: Dat is duidelijk en als ik het dus goed begrijp wil je dat er op datagebied niet veel gebeurt als we kijken naar onze eigen folders?

Ramon: Ik denk dat we hier al voldoende hebben, maar ik geloof dat onze belangrijkste groeikansen liggen in de breedte en het aanbod. Daarom moeten we ons richten op het visualiseren van data van concurrenten. Dit kan via de retailers zelf gebeuren of via een samenwerking.

Daarnaast wil ik nog één punt benadrukken: we krijgen bij WePublish nog te vaak vragen over waarom iets niet werkt en het dashboard is hier een voorbeeld van. Dit moet ook verbeterd worden.

Nick: Begrijpelijk en hier heb ik goed nieuws. Uit eerdere interviews is gebleken dat Publitas misschien achterloopt op het gebied van dashboarding, maar waar ze in uitblinken vergeleken met WePublish is de architectuur, performance en schaalbaarheid. Als ze dit goed weten te combineren, zie ik geen probleem meer in de toekomst.

Addendum interview: Requirements deelvraag 5

Nick: Ha Ramon, fijn dat we nog even kunnen meeten. Zoals ik je op Slack had gestuurd had ik nog een paar korte aanvullende vragen betreft het toekomstige dashboard.

Ramon: Vertel, waarmee kan ik je helpen?

Nick: Omdat je één van de primaire stakeholders bent van het toekomstige dashboard heb ik dit even ingeschoten. Ik ben namelijk de requirements aan het opstellen, met andere woorden waaraan moet het dashboard voldoen en wat valt er buiten de scope.

Om mee te beginnen de vraag, wat moet er minimaal in het dashboard zitten en welke punten zijn volgens jou belangrijk?

Ramon: Wat betreft de technische aspecten heb ik niet veel kennis, daarvoor hebben we jou en Marc. Maar wat ik wel belangrijk vind voor het dashboard zijn een paar punten. De gebruikersvriendelijkheid staat voorop. Het dashboard moet helder en eenvoudig te begrijpen zijn,

zodat ook niet-technisch bezoekers met slechts een paar filters de benodigde gegevens kunnen opvragen.

Daarnaast vind ik het zeer belangrijk dat alleen data die voor hen bedoeld is zichtbaar is. Dus een ALDI mag absoluut geen data inzien van AH bijvoorbeeld. Oh en tot slot, Publitas opmaak. Dit is in het oude dashboard verouderd.

Nick: Deze punten zijn helder, heb je nog verdere punten die je belangrijk vindt?

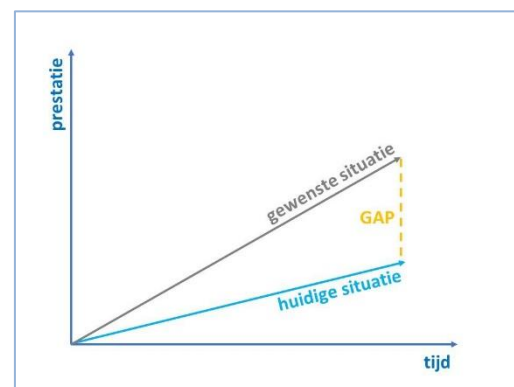
Ramon: Nee eigenlijk niet, qua techniek heb ik er geen verstand van en zolang het dashboard gebruikersvriendelijk is en er strak uit ziet ben ik al gauw tevreden.

Bijlage 12 – Gap-Analyse van Ist naar Soll

In deze GAP-analyse wordt de huidige staat en situatie van de data tracking in kaart gebracht van beide organisaties, ook wel de Ist-situatie genoemd in het GAP-model, vergeleken met de toekomstige gewenste situatie omtrent de data tracking, die in het GAP-model de Soll-situatie wordt genoemd (Gertjanschap, GAP-analyse vanuit Ist naar Soll - Managementmodellensite, 2023) .

Het verschil wordt vaak gemeten aan de hand van kwantitatieve metingen. In dit geval wordt er echter niet met kwantitatieve metingen gewerkt, maar wordt er gebruikgemaakt van informatie uit de data-mapping en de interviews met de primaire stakeholders. Het vaststellen van het verschil is echter pas de eerste stap. Het doel van het GAP-model is om op zoek te gaan naar de oorzaken van de ongewenste situatie en hierbij passende maatregelen te treffen om naar de Soll-situatie te komen.

De uitkomsten van het GAP-model zullen dienen als advies/aanbevelingen over hoe Publitas in de toekomst moet omgaan met het tracken van bezoekersdata. In grote lijnen loopt dit tussen de organisaties Publitas en WePublish al gelijk, echter op detailniveau verschilt het. Nu is het van belang om de juiste methode te kiezen. Om de correcte Soll-situatie te bepalen, zullen de stakeholder interviews als input dienen.



Figuur 43 Gap analyse

IST:

De huidige situatie is momenteel:

- Twee aparte folder viewers met een eigen back-end, die niet met elkaar samenwerken.
- Data van beide folder viewers worden apart gemeten door andere type data trackers.
 - De manier waarop en wanneer data wordt getrackt verschilt tussen de verschillende viewers.
- WePublish is zeer gericht op het gebruik van externe data in de vorm van consultancy en dashboarding. Publitas heeft alleen een basis dashboard.
- Opmaak van de dashboards zijn nog in eigen stijl, dus WePublish huisstijl en Publitas huisstijl.
- Geen tot minimale kennisdeling en een gebrek aan een gezamenlijke data-strategie.

SOLL:

De gewenste situatie

- WePublish-klienten gemigreerd naar de Publitas-back-end, waardoor er één uniforme back-end is voor de folders op de eigen platformen.
- Uniforme data-tracking met behulp van Snowplow.
 - Pageviews dienen gemeten te worden op dezelfde wijze als bij WePublish: pageviews worden alleen geregistreerd bij het doorklikken van pagina's en niet wanneer een product wordt aangeklikt.
- Gezamenlijke datastrategie omtrent dashboarding & consultancy.
 - Basis dashboard van Publitas en een geavanceerd dashboard van WePublish.
- Dashboarding dient de huisstijl van Publitas te hanteren.



publitas