

# Plan van aanpak

## Ontwikkel flexibele data-aggregaties voor Shop2market

door

S. Oldeman

ter verkrijging van de graad van Bachelor HBO-ICT  
aan de Hoge school Utrecht, Institute for ICT.  
In het openbaar te verdedigen in Juni, 2016.



Student nummer:	1571564
Project duur:	7 maart 2016 – 31 mei 2016
Afstudeer examinatoren:	M. Dumont, HU, docentbegeleider
	J. W. Pauw, HU
	M. Jorissen, Shop2market



# Inhoudsopgave

<b>1</b>	<b>Opdracht</b>	<b>1</b>
1.1	Achtergrond . . . . .	1
1.2	Aanleiding . . . . .	2
1.3	Organisatie . . . . .	2
1.4	De kwestie . . . . .	3
1.5	Doelstelling . . . . .	3
1.6	Type opdracht . . . . .	3
<b>2</b>	<b>Onderzoeksplan</b>	<b>5</b>
2.1	Onderzoeksvragen . . . . .	5
2.2	Literatuur . . . . .	5
2.3	Onderzoek methode . . . . .	6
<b>3</b>	<b>Planning</b>	<b>7</b>
3.1	De opdracht . . . . .	7
3.2	Activiteiten . . . . .	7
3.3	Resultaten . . . . .	8
3.3.1	Oplevering . . . . .	8
3.4	Scope . . . . .	9
3.5	Randvoorwaarden . . . . .	9
3.6	Risico's . . . . .	10
3.7	De projectorganisatie . . . . .	11
3.7.1	Werkwijze en methodieken . . . . .	11
3.7.2	Betrokkenen . . . . .	11
3.7.3	Werkwijze en verantwoordelijkheid student . . . . .	11
3.8	Ethische afweging . . . . .	11
<b>A</b>	<b>Overzicht planning en fases</b>	<b>13</b>



# 1

## Opdracht

In dit hoofdstuk wordt beschreven bij welke organisatie de opdracht zich afspeelt. Wat waren de ontwikkelingen binnen de organisatie en wat zijn de redenen om een nieuw project te starten?

### 1.1. Achtergrond

Shop2market is een software development bedrijf in de business-to-business sector. Het bedrijf heeft een missie om organisaties te helpen de winst uit online advertentie- campagnes te maximaliseren. Voor alsnog diende shop2market als een IT oplossing ondersteunend aan het adviesbedrijf. Een grootte hoeveelheid van deze klanten waren webwinkels in het A segment. Maar omdat de integratie met een webwinkel vaak maatwerk opleverde, duurde een integratie gemiddeld zes tot acht maanden. Hieruit valt ook te concluderen dat veel bedrijven niet de technologische middelen in huis hebben om zelfstandig te kunnen starten met adverteren.

Daarom werd in begin 2015 gestart met de ontwikkeling van een nieuwe dienst: Adcurve. Met alle ervaring vanuit de adviesorganisatie zijn veel processen vertaald naar functionaliteiten. Door de diverse functionaliteiten<sup>1</sup> in Adcurve kan de webwinkелеigenaar zijn online marketing campagnes controleren en binnen budget houden. Dit is mogelijk doordat Adcurve een partij is tussen de webwinkel en publishers. Door platform zoals SEOShop of Magento is het mogelijk webwinkels binnen enkele minuten te integreren. Op basis van verzamelde gegevens zoals afkomstige bezoeken, bestellingen en advertentiekosten worden de nodige statistieken berekend. Met deze gegevens wordt de winstgevendheid per advertentie berekend.

Dit alles heeft als gevolg dat Shop2market op dit moment webwinkels in het midden en klein bedrijf bedient, maar internationaal op een veel groter volume. Webwinkels zijn nu binnen enkele minuten geïnstalleerd en kunnen hun producten gemakkelijk adverteren via zogeheten publishers. Publishers zijn de bedrijven die de advertenties publiceren. Het soort advertenties verschillen nogal per publisher. Denk bijvoorbeeld aan ingekochte zoekresultaten, producten op prijsvergelijkers of affiliaties, maar ook producten op marktplaatsen. De gebruiker kan zelf publishers installeren binnen Adcurve zodat de benodigde integratie automatisch wordt afgehandeld.

---

<sup>1</sup> Denk hierbij aan datavisualisaties en beheeracties, soms ook wel "Actionable insights" genoemd.

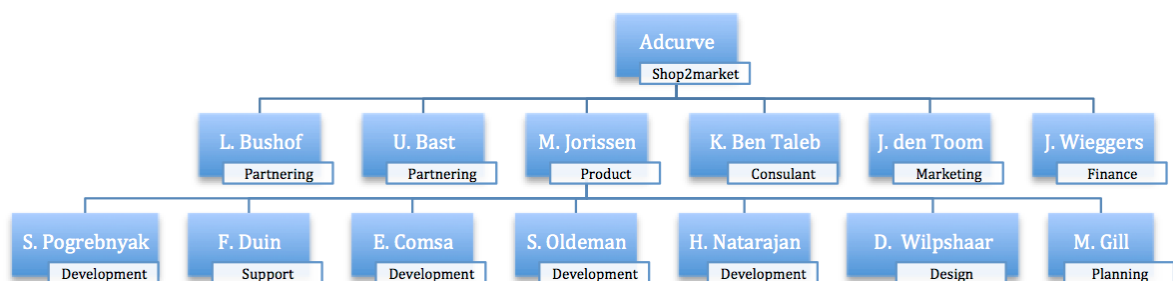
## 1.2. Aanleiding

Het afgelopen maanden zijn er vijf publishers geïntegreerd waarvoor een volledige dataintegratie plaatsvindt. Bij deze publishers worden kosten niet door ons berekend op basis van een percentage of vast bedrag. De kosten die in rekening zijn gebracht worden gerapporteerd en door Adcurve geïmporteerd. Omdat de belangrijkste functionaliteiten berusten op de beschikbaarheid van statistieken ligt dit proces aan de kern van de dienst. Helaas verloopt het importeren niet altijd zonder fouten. Het actief bewaken van datakwaliteit is hierdoor een prioriteit geworden. In de huidige situatie is dit nog lastig, doordat het berekenen van de statistieken een langdurig proces is.

De huidige strategie is om meer klanten aan te trekken door meer landen en industrieën te kunnen ondersteunen. Het is te verwachten dat het aantal data integraties met publishers zal toenemen, en het probleem daardoor groter wordt. Daarnaast is de huidige situatie niet optimaal om mee te beginnen. De statistieken zijn s'middags pas beschikbaar, waardoor veel functionaliteiten werken met data van meer dan een dag oud. Er is hierdoor een toenemende wens ontstaan om de huidige oplossing te herzien. Dit moet internationale groei van Adcurve ondersteunen, en ruimte bieden om functionaliteiten betrouwbaarder en te maken.

## 1.3. Organisatie

Shop2market is met zijn team gevestigd in Hilversum en kent op dit moment achttien werknemers (zie figuur 1.1). De organisatie kan naar de theorie van (Mintzberg, 1983) worden omschreven als een Adhocracy: "Door de innovatieve aard van projecten is een organisatie gebaat bij flexibiliteit. Een formele hiërarchische structuur werkt daardoor minder goed." Dit is herkenbaar en valt terug te leiden naar de professionele houding die van werknemers wordt verwacht. Er wordt autonomie gegeven om zelf structuur aan te brengen wanneer dit nodig is.



Figuur 1.1: Organogram waarin het team en de relaties binnen Shop2market worden afgebeeld.

## 1.4. De kwestie

Zoals te lezen valt in de de aanleiding zijn er meerdere redenen voor dit project.

1. De tijd die het kost om statistieken te berekenen is aan de hoge kant. De berekeningen worden uitgevoerd met behulp van MongoDB MapReduce. Het team voorziet dat deze technologie niet genoeg schaalbaarheid biedt. Dit omdat rekentijd linear is toegenomen in relatie tot de hoeveelheid data. Met de verwachte groei van Adcurve komen nieuwe requirements aan het licht en er moet naar een nieuwe oplossing worden gezocht.
2. Bij voorkeur worden publishers geïntegreerd met behulp van API's oftewel; een externe databron. Op deze manier worden berekeningen uitgevoerd met preciese advertentiekosten. Maar doordat externe factoren nu een rol spelen in de berekeningen is het niet te garanderen dat de uitkomst altijd correct is. Zodra fouten intern of extern hersteld zijn, worden berekeningen voor een bepaalde dag, webwinkel of publisher opnieuw uitgevoerd. Het uitvoeren van dit soort correcties is tijdrovend door de huidige implementatie en gebruikte technieken.
3. Als laatste ontstaat er een kans door de huidige problematiek op te lossen. Webwinkels ontvangen tot soms tot 30 dagen na een bestelling een retournering van een of meerdere producten. Dit betekent dat een product niet verkocht is en de omzet uit de bestelling lager ligt dan is berekend. Het is wenselijk om berekende statistieken met betrekken tot retour bestellingen opnieuw te kunnen berekenen.

## 1.5. Doelstelling

In het kort moeten webwinkelenaren in staat zijn om beslissingen te maken op basis van correcte en actuele gegevens in Adcurve. Dit betekent dat de gegevens die Adcurve toont altijd te verklaren zijn en overeenkomen met de werkelijkheid. Als voorbeeld hiervan zijn gegevens over de vorige dag voor kantooruren beschikbaar en worden retouropdrachten ook verwerkt in Adcurve. Als laatste moeten gegevens accuraat en betrouwbaar zijn, tenzij anders vermeld staat. Daarom moet tijdig herstel van fouten mogelijk zijn.

## 1.6. Type opdracht

Aan de hand van beschikbare technieken wordt er gekozen een aantal mogelijke oplossingen te proberen middels een Proof of Concept. Dit sluit goed aan bij de aard van een onderzoeksopdracht. Door dit onderzoek moet duidelijk gaan worden: Hoe kan de opdracht worden opgelost? Is dit mogelijk? En wat is er nodig om de oplossing naar productie te krijgen? Door de opdrachtgever is gevraagd om een aantal Proof of concept 's te ontwikkelen om tot een inzicht over de oplossing te komen. Er is hierdoor sprake van een ontwikkel opdracht.





# 2

## Onderzoeksplan

In de kwestie zijn de problemen - beter uitdagingen te noemen - omschreven. Omdat een oplossing niet voor de hand ligt, wordt er een onderzoek opgesteld. De leidende hoofdvraag wordt in dit hoofdstuk helder.

### 2.1. Onderzoeksvragen

Gedurende de uitvoering van de opdracht wordt de volgende hoofdvraag beantwoord:

*"Hoe verzorgt een nieuwe implementatie voor het up-to-date houden van statistieken in Adcurve zodat gegevens altijd te verklaren zijn?"*

Om de hoofdvraag te beantwoorden zijn de volgende deelvragen geformuleerd:

1. Wat zijn de functionele en niet functionele eisen waaraan de oplossing moet voldoen?
2. Wat zijn toonaangevende methodes en technologieën om statistieken te berekenen, die zowel passen bij de wensen en eisen van de opdracht?
  - (a) Hoe kan de opdracht worden opgelost?
3. Welke scenario's komen met regelmaat voor waardoor gegevens niet te verklaren zijn?
  - (a) Wat zijn de gerelateerde technische factoren waardoor de scenario's voor verandering zorgen in de huidige situatie?
  - (b) Wat zijn de mogelijke strategieën en technieken om dit op te lossen?
4. Wat zijn de gepresenteerde oplossingen en waarom zijn deze volledig of niet?

### 2.2. Literatuur

Voorbeelden van mogelijk jargon die zullen voorkomen in de thesis zijn data aggregaties, data transformaties en het selecteren en installeren van big data tools. Deze concepten worden onderlegd door onder andere de volgende literatuur:

- Data Mining, Concepts and Techniques (Jiawei Han, 2001)
- I <3 Logs, Event data, stream processing, and data integration (Kreps, 2014)
- Fast Data Processing with Spark (Krishna Sankar, 2015)
- Real-Time Big Data Analytics (Barlow, 2013)

Tijdens de selectie wordt mogelijk gebruik gemaakt van verschillende fases uit "de Berenschot-methode" (Hans Cuppen, 2010). Daarnaast zal tijdens het voeren van gesprekken, interviews en presentaties binnen de organisatie mogelijk gebruik worden gemaakt van de theorie uit "Adviseer als tweede beroep, resultaat bereiken als adviseur" (Nathans, 2005).

### 2.3. Onderzoek methode

Voor kwalitatief onderzoek wordt een *Case studie* gebruikt om de problemen in de gegeven context te analyseren. Methodes binnen dit type onderzoek zijn: explanatory, descriptive en exploratory. (Dudovskiy, 2015). Het onderzoek is ontworpen om de fases van een Case study uit te voeren. Het ontwerp is omschreven in tabel 2.1.

Tabel 2.1: Onderzoek methodes met te gebruiken methoden/technieken/middelen per deelvraag

#	Deelvraag	Type vraag	Methode	Actie / Resultaat
1	Wat zijn de functionele en niet functionele eisen, waaraan de oplossing moet voldoen?	Descriptive	Interviews	MosCow prioriteiten lijst en checklist samenstellen
2	Wat zijn toonaangevende methodes en technologieën om statistieken te berekenen, en passen bij de wensen en eisen van de opdracht?	Descriptive	Literature-research	Analyseren van bronnen m.b.v. van checklist wordt een shortlist samengesteld
3	Welke scenario's komen met regelmaat voor waardoor gegevens niet te verklaren zijn?	Descriptive	Interviews, Literature-research	Inventariseren op te lossen scenario 's met prioriteit d.m.v. Impact analysis
3a	Wat zijn de gerelateerde technische factoren waardoor de scenario's voor verhindering zorgen in de huidige situatie?	Descriptive	Literature-research	Vergelijkingstabel huidige en wenselijke situatie met daarbij de technissche afhankelijkheden om een scenario te kunnen voorkomen
3b	Wat zijn de mogelijke strategieën en technieken om dit op te lossen?	Designing	Interviews, Literature-research	Door het toepassen van de vergelijkingstabel met gevonden technologieën worden mogelijke oplossingen ontworpen
4	Wat zijn de mogelijke oplossingen en hoe wordt dit gevalideerd?	Comparative	Literature-research, group-discussion	Door het Analyseren van mogelijke ontwerpen en groep discussie worden er een beperkt aantal Proof of concept's geformuleerd met eisen.
5	Wat zijn de gepresenteerde oplossingen en waarom zijn deze volledig of niet?	Explanatory	Case study	Conclusies op basis van van verzamelde gegevens tijdens Proof of concept- fase, zoals bijv. performance tests. Tabel van oplossingen met theorieën die het resultaat verklaren.

# 3

## Planning

In de planning wordt omschreven hoe de besluitvorming zal verlopen. Er is gekozen om met Agile development methodieken te werken. Daarnaast wordt verder toegelicht hoe de opdracht verder is afgebakend, zodat de student en opdrachtgever alle verwachtingen helder hebben.

### 3.1. De opdracht

Shop2market heeft altijd voor een uitdaging gestaan als het aan komt op het verwerken van data. Er zijn veel technologieën beschikbaar, maar niets is perfect. De grootte variatie aan use-cases maakt het kiezen van een oplossing geen voor de hand liggende keuze. Daarom is de opdracht gegeven om een Proof of concept te ontwikkelen om de risico's van een investering te minimaliseren.

### 3.2. Activiteiten

De project activiteiten zijn omschreven om de onderlinge samenhang en fasering te visualiseren. De fases hebben dezelfde indeling als de deelvragen uit het onderzoeksplan, zie sectie 2.3 op pagina 6. Per activiteit is een inschatting gegeven in uren. Het totaal aantal uren komt uit op 240 werkuren. Deze planning is uitgebreid terug te vinden in appendix A: Overzicht planning en fases.

1. Afleggen interviews rondom Requirements & Constraints, documenteren use-cases (4 uur)
2. Onderzoeken van toonaangevende, passende technologieën (24 uur)
3. Impact analyse uitvoeren rondom use-cases (4 uur)
  - (a) Onderzoek technische factoren die voor problemen zorgen (12 uur)
  - (b) Oplossing ontwerpen voor use-cases en Proof of concept (16 uur)
4. Bespreken van ontwerpen en evt. aanpassen van ontwerpen (2 uur)
  - (a) Vastleggen kwaliteitscriteria Proof of concept's (2 uur)
5. Uitvoeren van Proof of concept's
  - (a) Eerste iteratie (64 uur)
  - (b) Tweede iteratie (56 uur)
  - (c) Derde iteratie (56 uur)

De ingeschatte uren zijn uitsluitend de benodigde activiteiten voor het schrijven van de scriptie. Hier-voor wordt een inschatting gedaan van 200 uur. Dit geeft 40 uur voor het bijwerken van de scriptie bij ieder van de vijf fasen. Het aantal benodigde uren voor dit project komt hierdoor uit op 440 uur.

### 3.3. Resultaten

De volgende producten dienen in ieder geval te worden opgeleverd, deze worden omschreven met de kwaliteitscriteria die bekend zijn op dit moment. In de ontwerpfase van het project worden er kwaliteitscriteria gedocumenteerd per Proof of concept. Dit wordt voor iedere nieuwe iteratie gedaan, zie appendix A.

1. Impact Analyses van huidige probleem scenario's  
*Criteria: Tabel en/of visualisatie van scenario's, de waarschijnlijkheid en de technische en organisatorische impact hiervan*
2. Lijst van (non-)functional requirements  
*Criteria: Geprioritiseerd volgens MoSCoW methode*
3. Planning en Ontwerp per Proof of concept met kwaliteitscriteria  
*Criteria: Het ontwerp is besproken en goedgekeurd door de interne stakeholders. De iteratie is omschreven en ingevoerd in het ticketsysteem*
4. Programmatuur van Proof of concept's  
*Criteria: indien geslaagd, voldoet aan criteria van het ontwerp. De programmatuur is altijd vastgelegd in versiebeheer*
5. Test resultaten en bevindingen uit Proof Of concept's  
*Criteria: de vragen in het plan per Proof of concept zijn beantwoord*
6. De thesis met onderzoeksresultaten  
*Criteria: voldoet aan de eisen gesteld in de afstudeerleidraad, relevant aan de afstudeerperiode aan de Hogeschool Utrecht: HBO-ICT*

#### 3.3.1. Oplevering

Voor de organisatie is de uitvoer van Proof of concept's en het resultaat hiervan het eindproduct. Hiervoor zijn meerdere opleveringen gepland gedurende het project. Intern zijn deze ingedeeld in Agile iteraties van twee weken. Zie appendix A waarin alle interne oplever data staan vernoemd onder *demo*.

Voor de opleiding is het belangrijkste eindproduct de thesis, met de nadruk op het onderzoek. Deze wordt ingeleverd bij de Hogeschool Utrecht voor 31 mei 2016. Daarnaast worden er twee concept versies gecommuniceerd, gepland op 26 april 2016 en 10 mei 2016.

Concreet worden hierbij de volgende beroepsproducten aan het eind van het project opgeleverd:

- De thesis met alle onderzoeksresultaten.
- Resultaten van Proof of concept's, aldus de programmatuur.
- De bevindingen, te lezen in "Test resultaten en bevindingen uit Proof of concept's", zijn bij de thesis inbegrepen.

### 3.4. Scope

De statistieken worden momenteel berekend op een aantal dimensies: per jaar, per maand, per week en per dag. De aggregaties worden in een aantal stappen uitgevoerd, in de volgende volgorde:

- per geadverteerd product bij een publisher.
- totalen per publisher
- totalen voor de gehele webwinkel

Het berekenen van statistieken op het laagste niveau geeft de grootste uitdaging. Dit komt mede omdat de invoer data (input) een groter volume heeft op dit niveau, in vergelijking tot latere aggregaties. Ieder Proof of concept beperkt zich daarom tot het berekenen van een aantal metriecken op het laagste aggregatie niveau voor webwinkels: geadverteerde producten.

### 3.5. Randvoorwaarden

Voor dit project zijn er geen afhankelijkheden van derden partijen of andere projecten binnen in de organisatie. Daarom is de student slechts afhankelijk van de samenwerking van de organisatie. Omdat alle betrokkenen al langere tijd samenwerken, kunnen de volgende voorwaarden als vanzelfsprekend gezien worden:

- De organisatie houdt rekening met het project en de student in de reguliere planning van iteraties. De plannig zoals in apendix A is hiervoor het uitgangspunt.
- De stakeholders stellen tijd beschikbaar voor het afleggen van interviews en goedkeuring van documentatie.
- De student heeft toegang tot benodigde servers, data en eventuele lab omgeving.

### 3.6. Risico's

Gedurende het project zijn er een aantal momenten waarop de risico's worden geëvalueerd. De mogelijkheid bestaat om dit dagelijks te doen tijdens de *daily stand-up*. Zo kan er bijvoorbeeld tijdens de uitvoer van een Proof of concept worden gezien of de ingeslagen weg vitaal is. Hieronder worden een aantal mogelijke scenario's uitgeschreven en de mogelijke besluitvorming.

Tabel 3.1: Risico scenario's met mogelijke besluitvorming om de impact te minimaliseren

Scenario	Impact	Mogelijke onderneming
Tijdens een geplande iteratie / Proof of concept wordt duidelijk dat een ontwerp niet voldoet	Planning	Het huidige plan wordt verworpen. Er moet een nieuw ontwerp worden gemaakt. In de huidige iteratie worden nieuwe taken omschreven, ingeschat en uitgevoerd in ruil voor het vorige plan.
Een van de belangrijkste use-cases in het Proof of concept zijn naar tevredenheid opgelost, met uitzondering van enkele use-cases. Daarnaast is geconcludeerd dat niet alle use-cases door een techniek kan worden opgelost.	Kwaliteit	Indien er voldoende tijd over is zal er een mogelijk een iteratie worden gewijzigd of gepland. Hierin moet een oplossing worden ontworpen en ontwikkeld om de resterende use-cases op te lossen. Indien hier geen tijd voor is, is dit scenario een waardevolle uitkomst voor de opdrachtgever.
De schaalbaarheid van technologieën moet mogelijk worden bewezen, dit kan veel tijd in beslag nemen door het verplaatsen van data. Het uitvoeren van deze test moet geautomatiseerd kunnen.	Planning	Voor het geautomatiseerd testen is extra tijd nodig in de maand Mei, deze is momenteel niet gepland omdat het niet zeker is of deze activiteit vereist is. Daarnaast is het alleen van toepassing op geslaagde ontwerpen die in het Proof of concept naar voren zijn gekomen.

### 3.7. De projectorganisatie

De software ontwikkeling valt onder leiding van M. Jorissen. In nauwe samenwerking met M. Gill zijn zij verantwoordelijk voor de project analyse en planning. Het development team bestaat uit zes ontwikkelaars waaronder één ontwerper D. Wilpshaar. De architectuur wordt geleid door S. Pogrebnyak. Dit team is verantwoordelijk voor het uitwerken en oplossen van projecten. Hierbij wordt gebruik gemaakt van een aantal methodieken en gewoontes vanuit Agile-softwareontwikkeling, waaronder: Kanban, Daily standup, Retrospectives, Extreme Programming (XP), Test Driven Development (TDD) en Continuous integration (CI).

#### 3.7.1. Werkwijze en methodieken

Ondanks dat er in de werkwijze vaak een uitgebreide planning of requirements analyse ontbreekt, wordt de productiviteit en software kwaliteit toch bewaakt door de principes uit XP zoals TDD. Refactoring<sup>1</sup> speelt hier een grootte rol in. Hierdoor kan het team zich veroorloven om bij het ontwerpen en ontwikkelen van projecten minder vooruit te plannen. Software wordt hierdoor gevormd door requirements die in de huidige situatie aanwezig zijn. In tegenstelling tot het ontwikkelen met abstracties; zodat software flexibel genoeg is voor latere aanpassingen en requirements. Veel planning is hierdoor niet aanwezig en technische implementatie komt vooral tot stand door TDD en brainstorm sessies. Dit is mede mogelijk doordat de omgeving de mogelijkheid bied om fouten te maken.

#### 3.7.2. Betrokkenen

Betreft de stakeholders van het project zijn de volgende personen betrokken bij het project: Matthijs Jorissen (Opdrachtgever) en S. Pogrebnyak (technische eindverantwoordelijk). Daarnaast is Marco Dumont betrokken in zijn rol als docent-begeleider. S. Oldeman, de student heeft in dit project een dubbel rol als student en werknemer.

#### 3.7.3. Werkwijze en verantwoordelijkheid student

Het plannen en ontwikkelen van het software project valt onder de volledige verantwoordelijkheid van de student. De planning, doelstellingen en requirements worden in overleg met M. Jorissen vastgelegd. In verband met overdracht en kennis-deling zal er afstemming plaats vinden met S. Pogrebnyak over technische afwegingen. Dit gaat door middel van code reviews, discussies en document-deling.

### 3.8. Ethische afweging

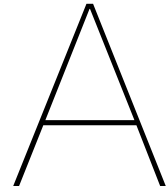
Er zijn kansen aanwezig om aspecten van het project ethische te evalueren. Dit moet echter nog verder worden onderzocht en is afhankelijk van het Proof of concept. Bijvoorbeeld doordat er gebruik wordt gemaakt van persoonlijke gegevens. Een andere mogelijkheid is dat het project op zichzelf wordt overwogen vanuit een ethisch perspectief. Bijvoorbeeld door te kijken of er ethisch is gehandeld tijdens het uitvoeren van het project. Wanneer er geen spraken is van een ethisch dilemma zal dit worden toegelicht.

---

<sup>1</sup>“Code refactoring is het proces waarbij de structuur van bestaande code wordt gewijzigd zonder de functionaliteit te wijzigen”...“dit komt ten goede aan de onderhoudbaarheid van code, en creëert nieuwe architectuur of object modellen om code makkelijker te kunnen uitbreiden met nieuwe functionaliteiten” (Martin Fowler, 2009)







## Overzicht planning en fases

In de volgende twee pagina's zijn de planning en fases in meer detail omschreven<sup>1</sup>. De onderzoeksmethode, zie tabel 2.1 op pagina 6, dient als basis voor deze planning.

In de "Breakdown structure, estimates per activity" wordt inzichtelijk gemaakt hoe de tijd is onderverdeeld per activiteit en Proof of concept. Er wordt rekening gehouden met de mogelijke risico's en daarom kan in werkelijkheid de uitvoering sneller dan hier ingepland.

In de lijst met lijst met activiteiten, zie sectie 3.2 op pagina 7, zijn de inschattingen in uren geschreven. Echter zijn de inschattingen origineel gedaan in dagen, omdat dit beter aansluit bij de werkwijze van Shop2market. Dit wordt dan ook gehanteerd in appendix .

In "Visualisation in calendar days based on estimates per activity" is te zien hoe de planning past binnen de iteraties bij Shop2market, aansluitend bij de werkwijze in de organisatie. Andere belangrijke momenten als inleverdata en concept versies van de thesis zijn hier ook in opgenomen.

---

<sup>1</sup> Deze pagina's zijn in het engels geschreven, omdat dit de voertaal is binnen het bedrijf

## Breakdown structure, estimates per activity

### Estimated activity in days

A.1. Requirements & Constraints: interview, documenting	0.5
A.2. Literature research, find matching technologies	3
A.3. Impact analysis on problematic scenario's	0.5
A.3.a Investigate technical factors which lead to problems	1.5
A.3.b Design solutions for scenaior's	2
B.4. Compare and Discuss designs, define POC's	1
C.5. Execute first POC	8
C.6. Execute second POC	7
C.7. Execute max third POC	7

***Average days needed for the project / research*** 28

### Estimated Proof of concept / Breakdown in days

Learning a tool and/or language	1.5
Installing / Getting started	0.5
Introduction in ecosystem libraries	2
transforming data	2
implementing the POC	3
possible debugging	1.5
performance testing	1.5
documentation	1

***Average Subtotal (incl. Risks)*** 8

## Visualisation in calendar days based on estimates per activity

iteration	period	mon	tue	wed	thu	fri
1	07-03 / 11-03					
	14-03 / 18-03	Plan				Demo
2	21-03 / 25-03					
	28-03 / 01-04	Plan				Demo
3	04-04 / 08-04					
	11-04 / 15-04	Plan		Dentist		Demo
4	18-04 / 22-04					
	25-04 / 29-04		Concept 1			
NA	02-05 / 06-05					
	09-05 / 13-05		Concept 2			
NA	16-05 / 20-05					Print Thesis
	23-05 / 27-05	End				

### Activities per iteration

Iteration 1	Activities A.1 to A.3
iteration 2	Activity B.4 and C.5
iteration 3	Activity B.4 and C.6
iteration 4	Activity B.4 to C.7

### Colors explained for when this document is printed in black/white:

	Light Blue
	Darker Blue

### Notes

Concept (1,2)	Concept versions of thesis are send to students' supervisor
Demo	The student will deliver / demo results from Proof of concept, the gathered data and observations
Dentist	The student has an appointment, personal note for reference
Plan	The stakeholders plan the next iteration
Print thesis	The thesis will be send to the copy shop
End	Planned date to deliver / send the thesis for Commission Final end date stated by Hogeschool Utrecht is on 31-mei-2016



# Bibliografie

- Mintzberg, H. (1983). *Structure in fives: designing effective organizations*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Jiawei Han, M. K. (2001). *Data mining: concepts and techniques*. San Francisco: Morgan Kaufmann Publishers.
- Nathans, H. (2005). *Adviseren als tweede beroep, resultaat bereiken als adviseur*. Deventer: Kluwer.
- Martin Fowler, J. F. (2009). *Refactoring: ruby edition*. Crawfordsville: Addison-Wesley.
- Hans Cuppen, B. C. (2010). *It in bedrijf, veranderingsanalyse, wfm en pakketselectie*. Den Haag, Netherlands: Sdu Uitgevers.
- Barlow, M. (2013). *Real-time big data analytics: emerging architecture*. Sebastopol, CA: O'Reilly Media Inc.
- Kreps, J. (2014). *I love logs*. Sebastopol, CA: O'Reilly Media Inc.
- Dudovskiy, J. (2015). Qualitative Research Case Studies. <http://research-methodology.net/research-methods/qualitative-research/case-studies/>. [Opgevraagd op 8-maart-2016].
- Krishna Sankar, H. K. (2015). *Fast data processing with spark*. Burmingham: PACKT publishing.