

# *Afstudeerscriptie*

*De ontwikkeling van de Fund Document Engine*

Hogeschool Utrecht  
Informatica  
Lennert Borra 1512226

FinFiles B.V.  
Wittevrouwensingel 100  
3514 AM Utrecht

Afstudeerscriptie Informatie

Hogeschool Utrecht  
Faculteit Natuur & Techniek  
Nijenoord 1  
3552 AS Utrecht

FinFiles B.V.  
Wittevrouwensingel 100  
3514 AM Utrecht

Utrecht, 2010

**Begeleiders**

Bedrijfsbegeleider: Dhr. W. Bijsterbosch  
Eerste examiner: ir. H. van de Ridder

**Auteur**

Lennert Borra  
lennert.borra@finfiles.nl  
Student Informatica 2006-2011

## VOORWOORD

Deze scriptie is geschreven ter afsluiting van de voltijd opleiding Informatica aan de Hogeschool Utrecht.

Gedurende vijf jaar ben ik bezig geweest met het succesvol afronden van de opleiding Informatica. Het eerste jaar verliep zonder enige problemen. De daarop volgende jaren kampte ik met motivatieproblemen. Gedurende de minor *E-Business* en de meeloopstage steeg mijn motivatie en heeft mij gebracht waar ik nu ben. Tijdens deze perioden ben ik erachter gekomen dat er meer komt kijken bij de ontwikkeling van applicaties dan alleen programmeren.

Om een kroon te zetten op mijn studie heb ik in opdracht van FinFiles B.V. een innovatief systeem ontwikkeld. Hiervoor wil ik een aantal mensen bedanken. Allereerst Jasper Staring voor het aanbieden van de uitdagende afstudeeropdracht. Daarnaast wil ik mijn bedrijfsbegeleider Warner Bijsterbosch bedanken voor de ondersteuning bij de ontwikkeling van de *Fund Document Engine*.

Ik bedank mijn afstudeerdocent Henk van de Ridder voor de begeleiding vanuit Hogeschool Utrecht.

Daarnaast gaat mijn dank uit naar alle collega's binnen FinFiles voor het creëren van een werkomgeving waar ik met alle plezier mijn afstudeeropdracht tot het gewenste resultaat heb gebracht. Mede dankzij jullie is het mogelijk gemaakt om mijn afstudeeropdracht succesvol af te ronden.

Tot slot rest mij nog ieder die geïnteresseerd is in deze scriptie veel leesplezier toegewenst.

*Lennert Borra*

*Utrecht, 10 december 2010*

## MANAGEMENTSAMENVATTING

FinFiles B.V. beheert en distribueert data voor de financiële industrie. FinFiles streeft ernaar om data betrouwbaarder, sneller en goedkoper aan te bieden. De aangeboden data wordt ontsloten in diverse innovatieve applicaties. FinFiles heeft voor het beheren van de financiële data twee systemen, waarbij op alle twee de systemen weinig tot geen kwaliteitscontrole kan worden uitgevoerd. FinFiles heeft door de slechte kwaliteitscontrole een marginaal inzicht in de actualiteit van de data.

De systemen bevatten enerzijds *kwalitatieve fondsdata*, waarbij men moet denken aan bijvoorbeeld maand-, jaarverslagen in de vorm van PDF documentatie. En anderzijds *kwantitatieve fondsdata*, bestaand uit dagelijkse fondsgegevens, welke wordt opgeslagen in een database. FinFiles is voor de kwantitatieve fondsdata afhankelijk van de levering door externe partijen.

Omdat op zowel de kwalitatieve als kwantitatieve fondsdata weinig tot geen kwaliteitscontrole kan worden uitgeoefend, is FinFiles gestart met de ontwikkeling en implementatie van een nieuw, allesomvattend systeem, het *Open Fund Source*. Het doel van het Open Fund Source: het op één plek beheren van de kwalitatieve en kwantitatieve fondsdata, waardoor de kwaliteit van de beschikbare data naar een hoger niveau wordt getild.

Om de onderhoudbaarheid en kwaliteit van de kwantitatieve fondsdata beter te kunnen garanderen heeft FinFiles besloten om de data door de *fondsaanbieders* te laten aanleveren in plaats van de externe partijen. FinFiles ontvangt in deze situatie de data vanaf de bron. Dit heeft als positief gevolg dat de fondsdata betrouwbaarder is. De fondsaanbieders dienen de data door middel van een *XML* bestand aan te bieden en zijn daarbij tevens eindverantwoordelijke. Voor bepaalde soorten data, zoals bijvoorbeeld benchmarks blijven echter altijd derde partijen nodig.

De kwalitatieve fondsdata, waarvan verwijzingen naar de PDF documenten worden opgeslagen, wordt op het moment handmatig verwerkt en beheerd. Onder andere hierdoor valt er nagenoeg geen controle uit te voeren op de actualiteit van de fondsdata. Het is de taak van de afstudeerstudent geweest om het beheer van deze kwalitatieve fondsdata naar een hoger niveau te brengen. FinFiles heeft voor dit systeem één belangrijke eis: het kunnen opbouwen van een historie met betrekking tot de kwalitatieve fondsdata.

Om de actualiteit te kunnen garanderen is er voor gekozen om de verantwoording van de aanlevering van de kwalitatieve fondsdata ook bij de fondsaanbieders te leggen. Hiervoor is een apart XML formaat opgesteld. De fondsaanbieder heeft de keuze om het XML bestand aan te leveren vanaf een eigen locatie of, indien dit niet mogelijk is, te laten hosten door FinFiles. Het XML bestand bevat de benodigde fondsgegevens plus de verwijzingen naar de desbetreffende PDF documenten.

Om de processen te automatiseren is er een systeem ontwikkeld, genaamd Fund Document Engine, die de XML gegevens uitleest en opslaat in een ontwikkelde database. Deze database wordt tevens gebruikt voor kwantitatieve fondsdata. De verwijzingen naar de PDF documenten worden verwerkt door de Fund Document Engine die de bestanden opslaat op een FinFiles server. Om de kwaliteit en actualiteit van de opgeslagen fondsdata te garanderen worden er tijdens de verwerking van de XML bestanden diverse kwaliteitscontroles uitgevoerd.

Om de kwaliteitscontroles te monitoren is er een *dashboard* ontwikkeld. Dit dashboard monitort de verwerking van de aangeleverde XML bestanden. Voor het ontlasten van de server, waar de Fund Document Engine op gaat draaien, kan met behulp van het dashboard de verwerking van de diverse XML bestanden periodiek worden gemanaged door de datamanager.

Het automatiseren van de processen zorgt voor een verbetering van de kwaliteit van de gedistribueerde data en biedt FinFiles nieuwe perspectieven voor de toekomst.

## INHOUDSOPGAVE

Voorwoord .....	3
Managementsamenvatting .....	4
Inhoudsopgave .....	5
1. Inleiding .....	8
2. FinFiles .....	9
2.1 FinFiles een uitdagende organisatie .....	9
2.2 Positie student binnen FinFiles .....	10
3. FinFiles financieel data specialist .....	11
3.1 Kwalitatieve en kwantitatieve fondsdata .....	11
3.2 Kwaliteitsprobleem kwalitatieve fondsdata .....	11
3. Oplossing: het Open Fund Source .....	13
3.1 Uitgangssituatie .....	13
4. Fund Document Engine .....	14
4.1 Verwerkingsengine .....	14
4.2 Dashboard .....	14
4.3 Te realiseren functionaliteiten .....	16
5. Succesfactoren .....	18
5.1 Kansen .....	18
5.1.1 XML .....	18
5.1.2 ABN AMRO .....	18
5.1.3 Fondsnieuws.nl .....	18
5.2 Bedreigingen .....	19
5.2.1 Beleid fondsaanbieders .....	19
5.2.2 Kennis van distributie .....	19
6. Opgeleverde producten .....	20
6.1 Documentatie .....	20
6.1.1 Onderzoeksrapport aanlevermethoden .....	20

6.1.2 Functioneel ontwerp Fund Document Engine.....	20
6.1.3 Technisch ontwerp Fund Document Engine.....	20
6.1.4 Testrapport.....	21
6.1.5 Gebruikershandleiding.....	21
6.2 Programmatuur .....	22
6.2.1 Verwerkingsengine .....	22
6.2.2 Dashboard .....	22
7. Gebruikte technologieen .....	23
7.1 Projectmethodiek.....	23
7.2 Gebruikte programmeertalen en technieken.....	23
7.2.1 XML en XML schema.....	23
7.2.2 PHP .....	24
7.2.3 MySQL.....	24
7.2.4 MVC-model.....	24
7.2.5 Smarty .....	24
8.Ontwikkeling Verwerkingsengine.....	25
8.1 Aanlevermethode.....	25
8.1.1 Opstellen XML formaat.....	25
8.1.2 Validatie aan de hand van XML schema.....	26
8.2 Verwerkingsengine .....	27
8.2.1 GetFile.....	27
8.3.2 DownloadDocument .....	28
9. Ontwikkeling dashboard .....	29
9.1 Manage cronjob.....	29
9.2 Manage xml location .....	29
9.3 Documentprocessinglog.....	30
9.4 XMLprocessinglog .....	30
9.5 Statistieken .....	31

10. Testing .....	32
11. Implementatie .....	33
11.1 Technische implementatie .....	33
12. Conclusie en aanbeveling .....	34
13. Evaluatie.....	35
14. Verklarende woordenlijst .....	36
15. Bronvermelding .....	38
15.1 Internetbronnen .....	38
15.2 Literatuur .....	38
16 Bijlagen.....	39
Bijlage A: Databasemodel.....	39
Bijlage B: Gebruikte methoden .....	40
Bijlage B1: GetFile.....	40
Bijlage B2: DownloadDocument .....	41
Bijlage C: Fragment funddocument.XSD .....	42
Bijlage D: Source structuur .....	43

## 1. INLEIDING

De scriptie die voor u ligt dient ter verantwoording van de door de afstudeerstudent gerealiseerde afstudeeropdracht en heeft als doelgroep lezers met een HBO-ICT achtergrond. De opdracht is uitgevoerd in opdracht van FinFiles B.V.

De scriptie geeft een duidelijk beeld van het probleem wat FinFiles heeft met enkele van zijn producten. De diverse hoofdstukken geven de lezer inzicht in het ontstaan van het probleem. De bijbehorende oplossing, welke is gerealiseerd door de student, wordt nader toegelicht. De gemaakte keuzes tijdens de realisatie van het project worden in de scriptie verantwoord. Tot slotte wordt het eindresultaat beschreven en er worden aanbevelingen gedaan ten behoeven van het eindproduct.

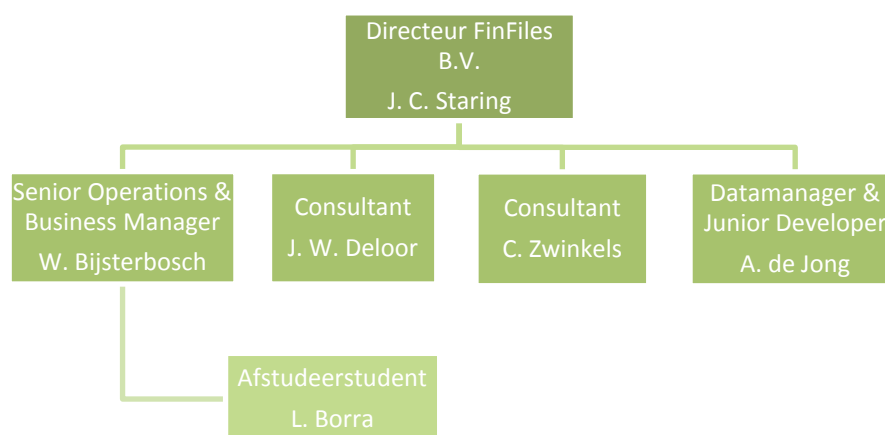


## 2. FINFILES

FinFiles B.V. is specialist op het gebied van data- en informatiemanagement voor de financiële industrie. Met de producten van FinFiles is informatie over financiële producten betrouwbaarder, sneller en goedkoper beschikbaar. De expertise is voor alle financiële dienstverleners van toegevoegde waarde. De nadruk wordt gelegd op de beleggingsindustrie, hierbij moet gedacht worden aan Asset Management, retail/private banking en vermogensbeheerders.

### 2.1 FINFILES EEN UITDAGENDE ORGANISATIE

FinFiles is een relatief kleine organisatie waar een lijnorganisatie wordt gehanteerd. Zoals het organogram (figuur 1) beschrijft bestaat FinFiles uit vijf fulltime medewerkers en één afstudeerstudent.



Figuur 1: Organogram FinFiles

De vijf medewerkers zijn verantwoordelijk voor een scala aan financiële applicaties. Het belangrijkste product van FinFiles is de *Fund Data Interface*, waarbij een zo compleet mogelijke dataset is ontsloten in een webapplicatie welke kwalitatieve en kwantitatieve financiële fondsgegevens bevat. De Fund Data Interface maakt het mogelijk om fondsgegevens op te vragen, te vergelijken en te publiceren in PDF formaat. De applicatie wordt ondersteund door video integratie.

Binnen de afdeling Operations & Business Development wordt momenteel gestart met de ontwikkeling en implementatie van een vooruitstrevende systeem voor *Asset Managers* en distributeurs van (informatie over) beleggingsproducten, genaamd Open Fund Source. De informatie die in deze toepassing wordt verzameld en gepubliceerd is grofweg in te delen in kwalitatieve en kwantitatieve fondsdata. FinFiles applicaties zullen op het gebied van datavoorziening gebruik gaan maken van de fondsdata die het Open Fund Source bevat.

Vanuit de WBSO <sup>1</sup> is voor 2010 8 maanden subsidie toegekend ter ondersteuning van de ontwikkeling van het Open Fund Source. Deze subsidiering wordt voortgezet in 2011.

<sup>1</sup> De WBSO, Wet Bevordering Speur- en Ontwikkelingswerk, is een fiscale stimuleringsregeling waarmee door de Nederlandse overheid een deel van de loonkosten voor ontwikkeling wordt gecompenseerd.

## 2.2 POSITIE STUDENT BINNEN FINFILES

Tijdens het afstudeerproject maakt de student deel uit van het Operations & Business Development team en zal verantwoordelijk zijn voor de kwalitatieve fondsdata binnen het Open Fund Source.

Voor de student is tijdens de afstudeerperiode één consultant beschikbaar voor inhoudelijke ondersteuning en begeleiding. Voor technische ondersteuning kan de student zijn bedrijfsbegeleider raadplegen. Daarnaast kan de student terecht bij de junior developer welke verantwoordelijk is voor de kwalitatieve data binnen de te ontwikkelen applicatie.

### 3. FINFILES FINANCIEEL DATA SPECIALIST

Naast de Fund Data Interface wordt de kwalitatieve en kwantitatieve fondsddata, welke FinFiles in zijn beheer heeft, in diverse applicaties ontsloten. Echter voldoen de huidige beheersystemen voor de fondsddata niet meer aan de hoge kwaliteitseisen die FinFiles aan zijn data stelt. Met name de betrouwbaarheid en actualiteit van de informatie spelen hierin een belangrijke rol. Het volgende hoofdstuk beschrijft de problemen ten aanzien van de kwalitatieve en kwantitatieve fondsddata.

#### 3.1 KWALITATIEVE EN KWANTITATIEVE FONDSDATA

De kwantitatieve fondsddata bestaat uit dagelijks fondsgegevens, dit zijn bijvoorbeeld NAVs, rendementen en risico's. Voor deze fondsddata is FinFiles momenteel afhankelijk van een *data feed* die ingekocht wordt bij een externe partij. Echter laat de kwaliteit van de aangeleverde data vaak te wensen over, wat resulteert in onbetrouwbare data binnen de applicaties.

De kwalitatieve fondsddata bestaat uit PDF documentatie die bijvoorbeeld maandelijks of jaarlijks worden uitgegeven door fondsaanbieders. Hierbij moet gedacht worden aan maandverslagen, prospectussen en bijvoorbeeld jaarverslagen. Deze kwalitatieve fondsddata wordt op het moment opgeslagen in het *Datasysteem*. Dit beheersysteem maakt het mogelijk om fondsaanbieder toe te voegen, te wijzigen en te verwijderen. Aan de fondsaanbieders binnen het Datasysteem kunnen fondsen worden gekoppeld met hun bijbehorende PDF documentatie, de kwalitatieve fondsddata. De opgenomen fondsen bevatten op hun beurt verwijzingen naar de PDF documenten. Deze verwijzingen linken door naar de externe sites van de fondsaanbieders. Het datasysteem bevat ruim 15.000 verwijzingen naar deze financiële PDF documenten. Het gebruik van externe verwijzingen brengt op de actualiteit van de fondsddata, mits goede controle, gevaar met zich mee.

De kwalitatieve en kwantitatieve fondsddata worden zoals eerder besproken ontsloten in diverse innovatieve applicaties. Het is voor FinFiles daarom van groot belang dat de fondsddata binnen de twee beheersystemen van hoge kwaliteit is.

#### 3.2 KWALITEITSPROBLEEM KWALITATIEVE FONDSDATA

De functionaliteiten binnen het Datasysteem zijn op het moment te beperkt om goede kwaliteitscontrole op de kwalitatieve fondsddata te kunnen uitoefenen.

De relatief slechte kwaliteitscontrole heeft diverse oorzaken. Het systeem is destijds ontwikkeld om steekproefsgewijs handmatige controle uit te oefenen op aangeboden data. In de loop van de tijd is de *database* uitgegroeid tot een beheersysteem met ruim 15.000 kwalitatieve (verwijzingen) documenten, waardoor handmatige controle nagenoeg niet meer haalbaar is.

Daarnaast is de slechte kwaliteitscontrole te wijten de diverse aanlevermethoden die FinFiles hanteert om het Datasysteem te voorzien van actuele fondsddata. FinFiles is afhankelijk van mailinglists en websites van de aanleverende partijen, de fondsaanbieders. Wanneer de aanleverende partij zijn data niet tijdig naar FinFiles communiceert betekent dit dat FinFiles geen actuele kwalitatieve fondsddata ter beschikking heeft. Met het Datasysteem heeft FinFiles geen inzicht in de aanlevermomenten van de fondsaanbieders. De datamanager houdt handmatig de updates met behulp van een Excel document bij, echter is deze niet altijd up-to-date.

De fondsdata welke de fondsaanbieders aanleveren worden in de vorm van links aangeleverd. De externe links worden door FinFiles opgeslagen in het Datasysteem. Hierdoor wordt het lastig controle uit te voeren op de actualiteit van de verwijzingen en nog belangrijker of de externe links nog bestaan. Zoals eerder beschreven bevat het datasysteem 15.000 bestanden en maakt het opslaan van externe links de controle nog lastiger. Een ander belangrijk aspect wat het opslaan van externe links met zich meebrengt, is het niet kunnen realiseren van een geschiedenis. FinFiles is in dit opzicht geheel afhankelijk van wat de aanleverende partij op dat specifieke moment aanbiedt. In de praktijk kan dit betekenen dat FinFiles 'dode' links gebruikt in de diverse applicaties. Door de PDF documenten achter de links op te slaan op kan FinFiles het probleem met de links verhelpen. Daarnaast brengt het als positief punt met zich mee dat er een geschiedenis van de documenten kan worden opgebouwd.

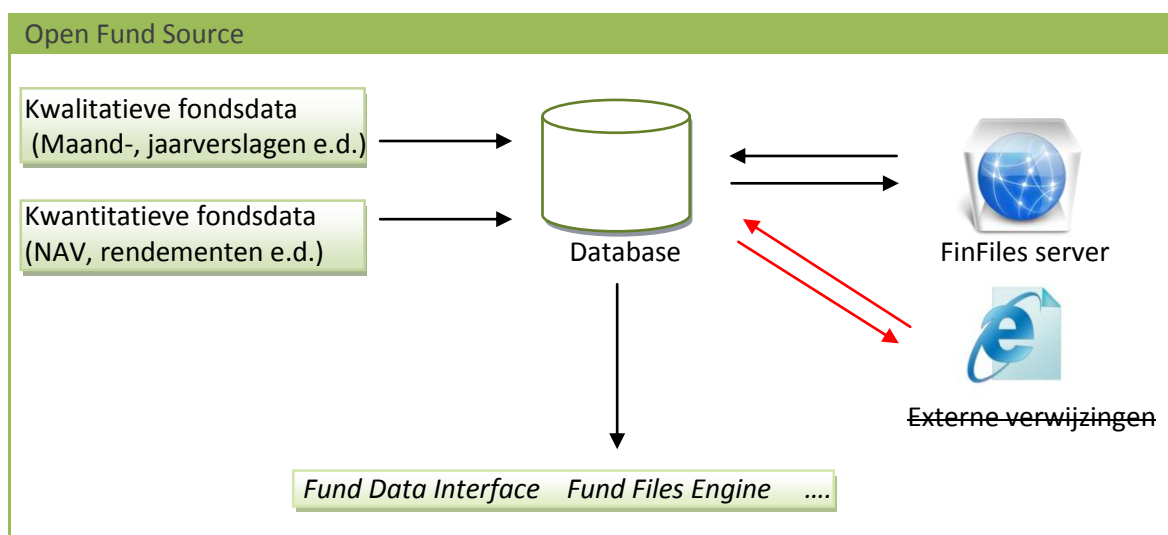
Tevens verschilt het moment van aanleveren door de diverse fondsaanbieders. Zonder specifiek aanlevermoment kan FinFiles niet controleren of zij de meest actuele kwalitatieve fondsdata beheren. Zoals hierboven beschreven wordt het Excel document voor het bij houden van de aanlevermomenten niet altijd bijgewerkt.

Kort gezegd zijn dit de belangrijkste redenen waarom FinFiles is gestart met de ontwikkeling van het Open Fund Source. Daarnaast vindt FinFiles het belangrijk dat het beheer van de kwalitatieve en kwantitatieve fondsdata binnen één systeem wordt onderverdeeld en ambieert FinFiles een systeem waarmee een geschiedenis kan worden opgebouwd. Hiermee kunnen completere datasets gebruikt worden binnen FinFiles applicaties.

### 3. OPLOSSING: HET OPEN FUND SOURCE

Zoals beschreven in de managementsamenvatting maakt de afstudeeropdracht deel uit van een groter innovatief project voor het beheer van kwalitatieve en kwantitatieve fondsdata, het Open Fund Source. Het Open Fund Source is in het leven geroepen om meer invloed te kunnen uitoefenen op de betrouwbaarheid en actualiteit van de kwalitatieve en kwantitatieve fondsdata.

Het project, Open Fund Source, is onderverdeeld in verschillende deelprojecten waarbij het de taak van afstudeerstudent is om het beheer en de inname van de kwalitatieve fondsdata naar een hoger niveau te tillen. In de praktijk betekent dit voor de student het ontwikkelen van een nieuw beheersysteem voor de kwalitatieve fondsdata, dat uiteindelijk ter vervanging van het huidige Datasysteem zal dienen. De nadruk zal binnen dit systeem, in tegenstelling tot de huidige processen, liggen op het automatiseren en het controleren van de documentinname waardoor er actuele fondsdata aan wordt geboden. Figuur 2 geeft de nieuwe situatie weer, waarbij in het rood de oude situatie wordt aangegeven.



Figuur 2: Het Open Fund Source

#### 3.1 UITGANGSSITUATIE

FinFiles heeft besloten om de gegevens binnen één database te ontsluiten. Er is in de beginfase van het project een database ontwerp opgesteld waarop de afstudeeropdracht moet worden gebouwd. Of het databasemodel in deze situatie voldoet aan de eisen die gesteld worden aan het afstudeerproject in het begin stadium nog de vraag.

Om het project te laten slagen gaat FinFiles uit van volledige medewerking van de fondsaanbieders. Het project gaat tijdens de ontwikkeling getest worden met diverse (grote) fondsaanbieders. In geval van succes zullen deze fondsaanbieders definitief hun data aanbieden aan FinFiles. Met de grote aangesloten partijen heeft FinFiles een goed argument om de overige gewenste partijen te kunnen aansluiten op het Open Fund Source.

## 4. FUND DOCUMENT ENGINE

De afstudeeropdracht, Fund Document Engine, is specifiek gericht op de automatisering van de inname en verwerking van de kwalitatieve fondsdata. De Fund Document Engine zal bestaan uit twee onderdelen, namelijk de *Verwerkingsengine* en het dashboard. De werking van de onderdelen worden in onderstaand hoofdstuk nader toegelicht.

### 4.1 VERWERKINGSENGINE

Het eerste onderdeel in de ontwikkeling van een systeem dat de verwerking van de kwalitatieve fondsdata op zich neemt is het automatiseren van de aanlevermethoden. Hierbij moet gekeken worden naar een uniforme aanlevermethoden waar alle fondsaanbieders zich aan dienen te houden. Om de processen zoveel mogelijk te automatiseren is er voor gekozen om de verantwoordelijkheid voor de aanlevering van de documenten bij de fondsaanbieder te leggen. De fondsaanbieders zijn erbij gebaat dat er actuele data aangeboden wordt en op deze manier hebben zij dit zelf in de hand.

Er is gekozen om een XML formaat op te stellen voor de data aanlevering. Een XML bestand kan gemakkelijk worden opgesteld en kan op een simpele wijze worden uitgelezen. Het voordeel van een XML bestand, boven bijvoorbeeld het aanleveren via een CSV bestand, is de validatie die kan worden uitgevoerd op het bestand voordat er gegevens opgeslagen worden in de database.

Nadat de gegevens uit het aangeleverde XML bestand zijn opgeslagen in de database, dient het systeem de verwijzingen naar de PDF documenten te verwerken. Tijdens de verwerking moet er duidelijk worden gecontroleerd of de opgeslagen verwijzingen bestaan. Zodra de verwijzingen correct zijn dienen de bestanden te worden opgeslagen op een FinFiles server. Zoals eerder beschreven streeft FinFiles er naar om een zo volledig mogelijke dataset aan te bieden en verwacht FinFiles dat er een historie wordt opgebouwd. De documenten op te slaan kan dit gemakkelijk worden gerealiseerd.

### 4.2 DASHBOARD

De Verwerkingsengine zal door middel van *cronjobs* worden aangeroepen. Een cronjob is een commando dat een script op een ingesteld tijdstip uitvoert. Doordat fondsaanbieders relatief veel XML bestanden gaan aanbieden, is er besloten om gebruik te maken van cronjobs. Deze maken het mogelijk om de aangeleverde XML bestanden van de fondsaanbieders op verschillende momenten te verwerken.

De aangeleverde XML bestanden bevatten fondsinformatie en verwijzingen naar de kwalitatieve documenten. Zoals eerder beschreven zijn er verschillende soorten PDF documenten zoals maand-, jaarverslagen en prospectussen. Een jaarverslag wordt bijvoorbeeld jaarlijks uitgeven. Met behulp van een cronjob kan een jaarverslag jaarlijks worden gecontroleerd en gedownload, dit bespaart aanzienlijk veel serverkracht en bandbreedte. Door de cronjobs slim in te regelen kunnen de verschillende XML bestanden en de daar bij behorende documenten op de juiste tijdstippen worden verwerkt. Het gebruik van cronjobs biedt hiervoor de beste oplossing.

Om de cronjobs te beheren moeten er beheerfunctionaliteiten worden ontwikkeld. Deze beheerfunctionaliteiten vallen onder het dashboard. Cronjobs moeten kunnen worden aangemaakt, gewijzigd en verwijderd.

Daarnaast moeten de XML locaties van de fondsaanbieders worden opgeslagen. Deze functionaliteiten worden ook in het dashboard opgenomen. Net als bij het cronjobbeheer worden de aanmaken, wijzigen en verwijderen functionaliteiten geïmplementeerd. De functionaliteiten kunnen op fondsaanbiederniveau worden gebruikt, dit wilt zeggen dat er per fondsaanbieder de cronjobs en XML locaties kunnen worden beheerd.

De belangrijkste functionaliteiten van het dashboard hebben betrekking op het loggen van de verwerkingsprocessen. Allereerst dient de verwerking van de XML gegevens te worden gelogd. Hierbij kan op aanbiederniveau gekeken worden welke XML bestanden wanneer verwerkt zijn en of deze zijn geslaagd, indien de verwerking niet geslaagd is wordt er gelogd wat er fout is gegaan.

Tevens wordt de verwerking van de verwijzingen naar de kwalitatieve fondsdata gelogd. Problemen die worden gelokaliseerd zullen veelal betrekking hebben op het herkennen van ongeldige links of bestanden, dit kan op fondsaanbiederniveau worden bekeken. Daarnaast kunnen de nog niet verwerkte verwijzingen bekeken worden. Het dashboard zal in eerste instantie worden gebruikt door de datamanager van FinFiles, maar dient dermate gebruiksvriendelijk te zijn zodat elke werknemer binnen FinFiles er gebruik van kan maken.

Wanneer nieuwe fondsaanbieders zich aanmelden om te worden opgenomen in het systeem moeten deze kunnen worden aangemaakt. Omdat dit geldt voor de kwalitatieve en kwantitatieve fondsdata staat het beheer van de fondsaanbieder los van het te realiseren dashboard. FinFiles heeft aangegeven deze functionaliteit als nice to have op te willen nemen, zodat er meer focus kan worden gelegd op de Verwerkingsengine en het lokaliseren van fouten.

### 4.3 TE REALISEREN FUNCTIONALITEITEN

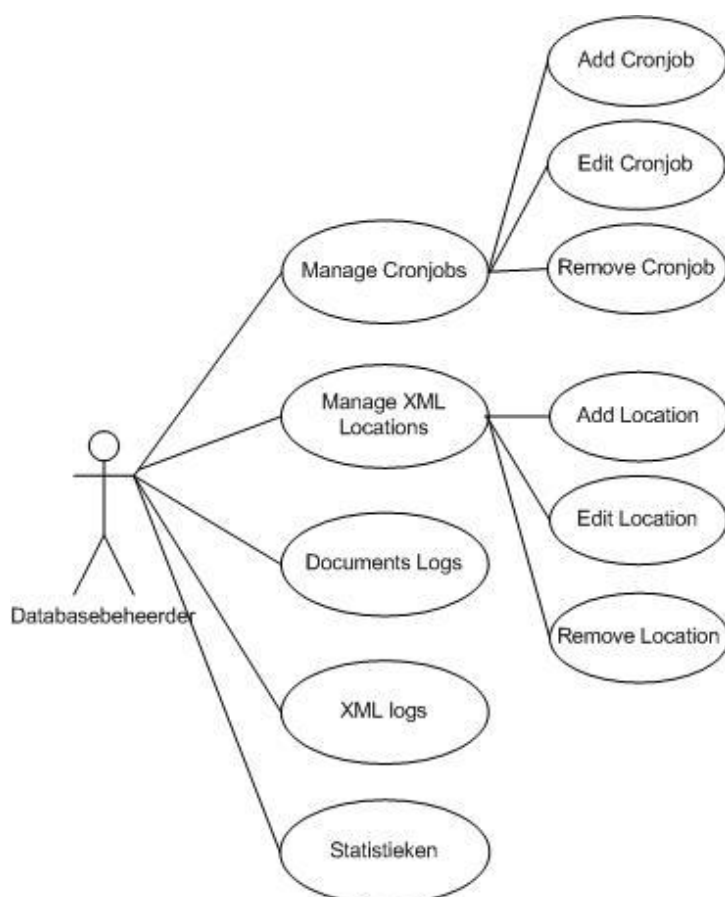
Samenvattend bestaat het gehele afstudeerproject uit de volgende functionaliteiten:

*Verwerkingsengine:*

- XML validatie aan de hand van een *XML schema*;
- Verwerking van de gegevens uit het aangeleverde XML bestand naar de Open Fund Source database;
- Loggen van verwerking van het aangeleverde XML bestand;
- Verwerking externe verwijzingen, betreft het opslaan van de kwalitatieve documenten op FinFiles server;
- Loggen van bestandsverwerking;

*Dashboard:*

Om de te ontwikkelende functionaliteiten van het dashboard overzichtelijk te weergeven, zijn de functionaliteiten weergegeven in het onderstaande *use case diagram* (figuur 3). Elke functionaliteit wordt kort toegelicht in de tabel op pagina 17.



Figuur 3: Use case diagram dashboard



Use case	Beschrijving
Manage cronjobs	Managecronjobs geeft de gebruiker inzicht in de bestaande cronjobs. Per fondsaanbieder kan worden bekeken welke cronjobs er zijn ingesteld. Voor de cronjobs kunnen drie acties worden uitgevoerd: add, edit en remove cronjob.
Add cronjob	Add cronjob geeft de gebruiker de mogelijkheid om nieuwe cronjobs toe te voegen. Naast de gewenste tijden kunnen er voor de cronjob diverse acties worden ingesteld. Deze acties hebben betrekking op de verwerking van het XML bestand en de bijbehorende kwalitatieve PDF documenten.
Edit cronjob	Wanneer cronjobs moeten worden gewijzigd kan dit met behulp van de edit functionaliteit.
Remove cronjob	Overbodige cronjobs kunnen gemakkelijk worden verwijderd met remove cronjob functie.
Manage XML locations	De diverse XML locaties kunnen worden bekeken met de manage XML locations functionaliteit. Evenals bij manage cronjobs kan dit op fondsaanbiederniveau worden bekeken. Vanuit manage XML locations kunnen er locaties worden toegevoegd, gewijzigd en verwijderd.
Add location	Add location biedt de gebruiker de mogelijkheid om voor een fondsaanbieder een XML locatie toe te voegen. Om rekening te houden met het HTTP en FTP protocol kan de gebruiker dit aangeven en indien nodig een gebruikersnaam en wachtwoord opgeven.
Edit location	Voor het wijzigen van een XML locatie is de functionaliteit edit location in het leven geroepen, doormiddel van een formulier kan een locatie gemakkelijk worden gewijzigd.
Remove location	Overbodige XML locaties kunnen worden verwijderd met remove location.
XMLprocessing log (XML logs)	De XML Logs loggen de fouten tijdens het importeren van de XML gegevens naar de database. De XML Logs geven een overzicht van de XML Locaties per fondsaanbieder. Er kan bekeken worden wanneer een XML locatie voor het laatst geprocesst is en of deze succesvol is. Indien mislukt worden de fouten getoond.
Documentprocessing log (Document logs)	Na de verwerking van een XML bestand worden de fondsgegevens opgeslagen in een tijdelijke tabel. De Documents Log functionaliteit laat de fondsen zien die geïmporteerd zijn maar waar de documenten nog niet van gedownload zijn. Dit kan op aanbiederniveau bekeken worden. Tevens kunnen de mislukte downloads per fondsaanbieder worden bekeken.
Statistieken	Om een zo duidelijk mogelijk beeld te kunnen vormen van de beschikbare data is er een statistieken pagina opgenomen. De statistieken pagina laat zien onder andere zien in hoeverre de datasets van de verschillende aanbieders compleet zijn.

## 5. SUCCESFACTOREN

Het is voor FinFiles van belang dat de implementatie van de Fund Document Engine succesvol verloopt. De huidige innamemethode vereist aanzienlijk veel tijd en er is onvoldoende grip op de kwaliteitscontrole. Het implementeren van het nieuwe systeem biedt voor FinFiles nieuwe kansen, maar ook na een succesvolle implementatie dient er rekening te worden gehouden met de bedreigingen. Dit hoofdstuk beschrijft de kansen en bedreigingen om het project naast een succesvolle oplevering ook te laten slagen in de markt.

### 5.1 KANSEN

#### 5.1.1 XML

In de huidige situatie leveren fondsaanbieders hun kwalitatieve fondsdata via de e-mail. Dit kan zijn in de vorm van verschillende bijlagen tot een lijst met de verwijzingen naar de diversen kwalitatieve documentatie. Het XML formaat wat is opgesteld is op een dermate simpele manier gestructureerd dat de drempel klein is om over te gaan op FinFiles XML formaat. Zodra fondsaanbieders geen gebruik willen gaan maken van de opgestelde XML standaard komt naast de Fund Document Engine ook het Open Fund Source in gevaar.

#### 5.1.2 ABN AMRO

FinFiles ontwikkelt in opdracht van ABN AMRO een nieuwe tool voor de beleggingsproducten op de website van ABN AMRO. De data ontsloten in deze tool wordt geleverd door 13 fondsaanbieders. Deze aanbieders zijn voor Nederlandse begrippen aanzienlijk groot in omvang ten opzichte van andere fondsaanbieders. FinFiles wilt voor het verzamelen van de data het Open Fund Source gebruiken. In de praktijk betekent dit dat de fondsaanbieders het nieuwe XML formaat dienen te gebruiken voor de aanlevering van de kwalitatieve fondsdata.

Met ABN AMRO en andere grote klanten is het voor fondsaanbieders aantrekkelijk om zich kunnen aan te sluiten op het Open Fund Source. Hun aangeleverde fondsdata wordt in dit geval gedistribueerd aan klanten van FinFiles. Voor de fondsaanbieder levert een kleine inspanning veel profijt op.

#### 5.1.3 FONDSNIEUWS.NL

Daarnaast wordt de data publiekelijk gedistribueerd door Fondsnieuws.nl. Fondsnieuws.nl is een initiatief van FD Mediagroep, dat gelanceerd is in augustus 2008. Het is een onafhankelijk journalistiek platform, dat zich richt op de snel aan belang winnende wereld van beleggingsfondsen en andere gestandaardiseerde beleggingsproducten.

Met grote klanten die de fondsdata gaan gebruiken is de stap kleiner voor fondsaanbieders om mee te werken aan de Fund Document Engine en het Open Fund Source.

## 5.2 BEDREIGINGEN

Naast de drivers om het gehele Open Fund Source te laten slagen, brengt de introductie van het Open Fund Source ook bedreigingen met zich mee.

---

### 5.2.1 BELEID FONDSAANBIEDERS

Met name binnen grote bedrijven is het niet eenvoudig om nieuwe processen te introduceren. Kleine wijzigingen in processen en workflows worden veelal door verschillende afdelingen beoordeeld en goedgekeurd of men werkt met een release-planning op het gebied van IT. Wat het accepteren van de introductie van een XML formaat kan bemoeilijken.

---

### 5.2.2 KENNIS VAN DISTRIBUTIE

In het huidige Datasysteem wordt naast de innamenmethode via de e-mail ook fondsdata verzameld door de databasemanager. In de praktijk houdt dit in dat de datamanager vanaf de diverse fondsaanbieder websites de verwijzingen naar de recentste documenten opzoekt en opslaat in het Datasysteem. De fondsaanbieder heeft hier echter geen weet van. Op deze manier heeft de fondsaanbieder geen kosten aan de aanlevering van de data om vervolgens toegevoegd te worden aan het Datasysteem. Deze fondsaanbieders moeten met de introductie van het Open Fund Source overtuigd worden van het voordeel van aanleveren via het opgestelde XML formaat.

## 6. OPGELEVERDE PRODUCTEN

Gedurende het project zijn er een aantal documenten en producten opgeleverd. Voor elk opgeleverde product wordt een uitgebreide beschrijving gegeven.

### 6.1 DOCUMENTATIE

De geschreven documenten zijn gebruikt ter documentatie voor de gerealiseerde producten. Daarnaast kunnen het technisch ontwerp en de gebruikershandleiding worden gebruikt voor doorontwikkeling van de Fund Document Engine.

#### 6.1.1 ONDERZOEKSRAPPORT AANLEVERMETHODEN

Om een beeld te kunnen vormen van de huidige manier van aanleveren van de kwalitatieve documentatie en in hoeverre dit kan worden geoptimaliseerd heeft de afstudeerstudent een onderzoeksrapport naar de aanlevermethoden geschreven.

In 2009 is het huidige datasysteem geanalyseerd op *usability* en op basis van het geschreven analyserapport is er een adviesrapport uitgebracht. Voor het onderzoeksrapport is gebruik gemaakt van de aanbevelingen uit het adviesrapport.

Het onderzoeksrapport beschrijft in detail de huidige manier van data-aanlevering, waarbij duidelijk de knelpunten worden benoemd. Daarnaast geeft het onderzoeksrapport mogelijke oplossingen weer om de processen te verbeteren. FinFiles heeft met de introductie van het Open Fund Source zelf al een duidelijk beeld van hoe de nieuwe inname van de kwalitatieve fondsdata georganiseerd dient te worden. Uiteindelijk is FinFiles' visie samen met de conclusies uit het onderzoeksrapport samengevoegd tot één methode waarmee de kwaliteit van de documentinname voldoet aan FinFiles' hoge kwaliteitseisen.

#### 6.1.2 FUNCTIONEEL ONTWERP FUND DOCUMENT ENGINE

In het functioneel ontwerp zijn alle functionele eisen die gesteld werden aan het project beschreven. De functionele eisen worden in het functioneel ontwerp kort toegelicht. Het functioneel ontwerp is gericht op de dashboardfunctionaliteiten. De diverse functionaliteiten zullen er uiteindelijk voor dienen om snel en gemakkelijk inzicht te krijgen in de fouten tijdens de diverse processen van de Verwerkingsengine zich hebben voorgedaan.

#### 6.1.3 TECHNISCH ONTWERP FUND DOCUMENT ENGINE

Het technisch ontwerp bevat de technische specificaties ten behoeve van de functionele eisen. Het technisch ontwerp gaat dieper in op de werking van de Verwerkingsengine. Er wordt nader beschreven welke methoden er zijn gebruikt en hoe de verschillende methoden onderling samenwerken.

De volgende aspecten worden in het technisch ontwerp behandeld:

- Ontwerp XML voor aanlevering kwalitatieve documenten;
- Gespecificeerd databaseontwerp gericht op de kwalitatieve data (bijlage A);
- Gebruikte programmeertaal inclusief verantwoording;
- Verantwoording gebruikte framework voor de realisatie van de front end;

Het technisch ontwerp is gebruikt voor de ontwikkeling van de Verwerkingsengine en kan door FinFiles worden gebruikt voor verdere ontwikkeling van het systeem.

---

#### 6.1.4 TESTRAPPORT

Om de kwaliteit van de aangeboden data te kunnen garanderen is er een testrapport opgesteld. Het testrapport richt zich afzonderlijk op respectievelijk de Verwerkingsengine en het dashboard.

Voor de Verwerkingsengine is het van belang dat er een slagingspercentage van 98% gehaald wordt. Dit wil zeggen dat indien een XML bestand verwerkt wordt, 98% van de kwalitatieve PDF documenten worden gedownload en opgeslagen op de desbetreffende FinFiles server. Voor de overige 2% zal FinFiles de fouten communiceren naar de fondsaanbieder, zodat de fondsaanbieders het XML bestand kunnen aanpassen.

Voor het dashboard zijn de verschillende use cases uit het functioneel ontwerp getest. Doormiddel van verschillende opgestelde testpaden wordt de usability van het dashboard getest.

Het testen van de Fund Document Engine is van belang omdat op deze manier van te voren kan worden gecontroleerd of de haalbaarheid van 98% rendabel is.

---

#### 6.1.5 GEBRUIKERSHANDLEIDING

Om het product succesvol over te kunnen dragen aan FinFiles is er een gebruikershandleiding opgesteld. De handleiding is geschreven voor medewerkers van FinFiles die in de toekomst het systeem gaan beheren. Tijdens het schrijven van de handleiding is er vanuit gegaan dat de gebruiker over enige technische kennis beschikt.

De gebruikershandleiding beschrijft de werking van de functionaliteiten binnen het dashboard. Aangezien de eindgebruiker beschikt over enige technische kennis is dit onderdeel een beknopt hoofdstuk binnen de handleiding.

Daarnaast zijn de verschillende methoden gebruikt voor de Verwerkingsengine gespecificeerd en is dit onderdeel uit het technisch ontwerp opgenomen in de handleiding. Dit om de eventuele doorontwikkeling van het systeem te bevorderen en kunnen eventuele fouten binnen de Verwerkingsengine aanzienlijk snel worden opgelost, zonder dat het technisch ontwerp wordt geraadpleegd.

## 6.2 PROGRAMMATUUR

De opgeleverde programmatuur bestaat uit de Verwerkingsengine en het bijbehorend dashboard.

### 6.2.1 VERWERKINGSENGINE

De Verwerkingsengine verwerkt, zoals eerder beschreven, de gegevens uit de aangeleverde XML en slaat uiteindelijk de kwalitatieve PDF documenten op op een FinFiles server. Voordat de Verwerkingsengine gerealiseerd werd is er een XML standaard opgesteld waar de fondsaanbieder zich aan moet houden. De XML standaard is op een dermate manier opgesteld dat het voor de fondsaanbieder, die in veel gevallen weinig technische kennis hebben, gemakkelijk in is te vullen. Het XML bestand wordt gedistribueerd naar de fondsaanbieders samen met een bijbehorend XML schema. Het XML schema wordt later tijdens de verwerking gebruikt voor de validatie van het XML bestand.

Omdat de Verwerkingsengine periodiek moet werken, wordt de Verwerkingsengine aangeroepen door diverse cronjobs. Hierdoor wordt het mogelijk gemaakt om verschillende kwalitatieve PDF documenten te downloaden wanneer nieuwe publicaties zijn gepubliceerd. Bijvoorbeeld door jaarverslagen jaarlijks in te plannen. Zodra alles correct is ingeregeld is er weinig beheer meer nodig.

Tijdens de verwerking van de XML bestanden worden alle mogelijke fouten gelogd door de Verwerkingsengine. Hierdoor kan er door de gebruiker overzicht gehouden worden op de kwaliteit van de aangeleverde fondsdata.

### 6.2.2 DASHBOARD

Om een zo duidelijk mogelijk inzicht te krijgen in de verwerking van de XML bestanden is er een dashboard ontwikkeld. Het dashboard geeft een overzicht in de verwerkte XML bestanden. De logs kunnen hier eenvoudig worden opgevraagd. Het dashboard biedt de gebruiker de mogelijkheid om de cronjobs te beheren en daarnaast kunnen de XML locaties worden gemanaged. De gerealiseerde functionaliteiten binnen het dashboard wijken af van de te realiseren functionaliteiten gespecificeerd in het afstudeervoorstel.

In het afstudeervoorstel zijn de beheerfunctionaliteiten ten behoeve van de fondsaanbieders opgenomen. Zoals beschreven vond FinFiles het van belang dat de Verwerkingsengine en het loggen van de gebeurtenissen de hoogste prioriteit kregen. Daarom is er tijdens de ontwikkeling besloten deze functionaliteiten achter wegen te laten.

In plaats daarvan is er een beheerfunctionaliteit voor de cronjobs en XML locaties ontwikkeld.

## 7. GEBRUIKTE TECHNOLOGIEËN

Het hoofdstuk “Gebruikte technologieën” beschrijft de technologieën die tijdens het afstudeerproject zijn gebruikt. De technieken lopen uiteen van projectmethodiek tot gebruikte programmeer technieken. De gebruikte technologieën worden in dit hoofdstuk verantwoord.

### 7.1 PROJECTMETHODIEK

In het projectvoorstel wordt gesuggereerd dat het project volgens de watervalmethode zal worden uitgevoerd. Er is destijds gekozen voor de watervalmethode omdat deze techniek bij de afstudeerstudent bekend is. Daarnaast legt de watervalmethode nadruk op de documentatie wat de onderhoudbaarheid van het systeem ten goede komt.

Tijdens de afstudeerperiode is er gebruik gemaakt van prototyping en is er afgeweken van het originele watervalmodel. Prototyping lijkt in veel opzichten op de watervalmethode. Door het gebruik van prototyping kon er tijdens de start van het project een beter beeld gevormd worden van het totale op te leveren product.

Door het prototyping model te gebruiken kon de afstudeerstudent, naast de op te leveren documenten, zich in een vroeg stadium richten op de realisatie van het project. Door kleine stukken code op te leveren en deze aan feedback te onderwerpen kon er een duidelijk beeld gevormd worden van het gehele project. Dit bracht als positief punt met zich mee dat er niet met verkeerde requirements gewerkt gaat worden in de bouwfase. Daarnaast heeft de opdrachtgever een beter beeld van het product dat wordt gemaakt.

### 7.2 GEBRUIKTE PROGRAMMEERTALEN EN TECHNIEKEN

Het volgende hoofdstuk besteedt aandacht aan de programmeertalen en technieken die gebruikt zijn bij de ontwikkeling van de Fund Document Engine.

#### 7.2.1 XML EN XML SCHEMA

Om de kwalitatieve fondsdata aan te kunnen leveren is er een XML standaard opgesteld. De XML standaard bevat alle data die nodig is om de kwalitatieve documenten gestructureerd op te slaan en te distribueren. Er is gekozen voor XML omdat XML gevalideerd kan worden aan de hand van een XML schema. Het XML schema beschrijft de structuur van het XML bestand, waardoor er voor de verwerking van het XML bestand al diverse validaties op elementen kunnen worden uitgevoerd.

Fondsaanbieders krijgen naast de XML standaard ook de beschikking krijgen over het XML schema. Hierdoor kunnen zij zelf het XML bestand valideren. Door strenge regels te hanteren zullen zich minder fouten voordoen tijdens de verwerking van het XML bestand.

### 7.2.2 PHP

Aangezien FinFiles voornamelijk internetapplicaties ontwikkelt in *PHP*, is het een logisch gevolg dat ook dit project in *PHP* is ontwikkeld. Deze eis werd vanuit FinFiles gesteld. Dit heeft alles te maken met de onderhoudbaarheid van het product na de oplevering. FinFiles heeft twee *PHP* specialisten in huis die verantwoordelijk zijn voor de ontwikkeling van de diverse producten.

Daarnaast is *PHP* de ideale taal voor het ontwikkelen van internetapplicaties. Het is een relatief eenvoudige taal waar achter een zeer grote community huist, waardoor een grote hoeveelheid aan tutorials en dergelijke beschikbaar is.

Als *IDE*, Intergrated Development Environment, is er gebruikt gemaakt van de Eclipse SDK met de *PHPEclipse* plug-in. Aangezien de afstudeerstudent hiermee bekend is.

### 7.2.3 MYSQL

De Fund Document Engine zal gaan draaien op een *MySQL* database. *MySQL* is een *Open Source* relationeel databasemanagementsysteem. Het wordt veelal gebruikt voor gegevensopslag voor websites en applicaties.

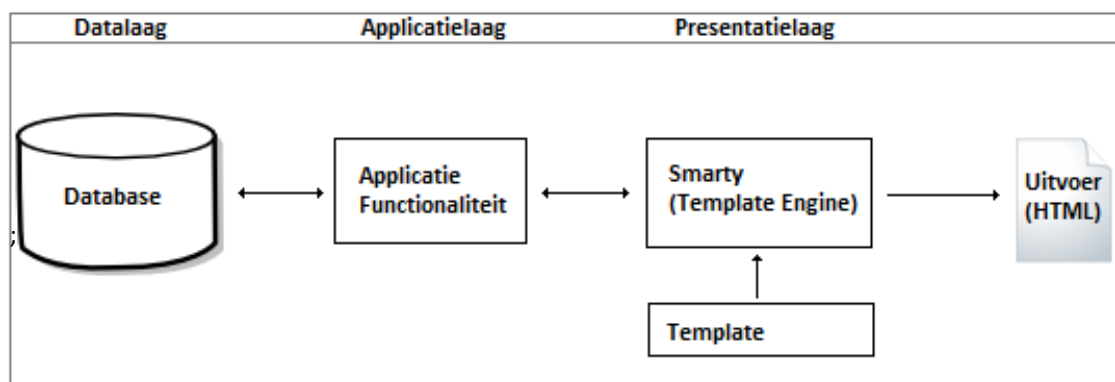
### 7.2.4 MVC-MODEL

Als ontwerppatroon is tijdens de ontwikkeling het MVC-model, Model-View-Crontroller, gehanteerd. Het MVC-model onderscheidt drie verschillende eenheden: model, view en controller. Waarbij respectievelijk het datamodel, de datapresentatie en de applicatie logica wordt worden gerepresenteerd.

Door de applicatie via het MVC-model in te richten kan deze in de toekomst goed onderhouden worden en kunnen nieuwe functionaliteiten gemakkelijk worden toegevoegd aan het systeem. De source-structuur van de Fund Document Engine wordt getoond in bijlage D.

### 7.2.5 SMARTY

Smarty is een *PHP framework* voor het scheiden van de presentatie (*HTML/CSS*) van de applicatielogica. Smarty genereert data vanuit de controller klassen en koppelt deze aan de desbetreffende template. Als de template wordt opgevraagd wordt de dynamisch toegevoegde data in de template (*HTML*) geplaatst. De gegenereerde *HTML* wordt op zijn beurt naar de client gestuurd. Figuur 4 representeert de algemene opzet van Smarty.



Figuur 4: Representatie algemene opzet Smarty



## 8. ONTWIKKELING VERWERKINGSENGINE

Het volgende hoofdstuk beschrijft de ontwikkeling van de Verwerkingsengine. De twee deelproducten worden uitvoerig beschreven. De verschillende stappen die de Verwerkingsengine tijdens het verwerken van het XML bestand doorloopt worden stapsgewijs beschreven. De functionaliteiten die het dashboard bevat worden, na de beschrijving van de Verwerkingsengine, aan de hand van screenshots toegelicht.

### 8.1 AANLEVERMETHODE

#### 8.1.1 OPSTELLEN XML FORMAAT

Het onderzoeksrapport heeft in detail de huidige aanlevermethoden van de kwalitatieve documentatie onder de loep genomen. Uit het rapport kwam naar voren dat de aanleverprocessen geautomatiseerd dienden te worden. FinFiles heeft destijds aangegeven dat de verantwoordelijkheid van de aanlevering van de data bij de fondshuizen wordt gelegd. Daarom is er gekozen om een XML formaat op te stellen waar alle fondsaanbieders zich aan moeten houden.

Voor het opstellen van het XML formaat moeten dezelfde attributen, welke het huidige Datasysteem bevat, worden opgenomen. Daarnaast is bij het opstellen van het XML formaat rekening gehouden met de complexiteit van het XML formaat, dit om fondsaanbieders niet te ontmoedigen. Figuur 5 geeft het XML formaat weer.

```
<?xml version="1.0" encoding="UTF-8"?>
<FundsDocumentXML version="1.0"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="FundsDocuments.xsd" >
  <CreationDate>2010-11-11</CreationDate>
  <Company>
    <Name>CompanyName</Name>
    <Funds>
      <Fund>
        <Name>FundName</Name>
        <Identifier>
          <Type>ISIN</Type>
          <Value>NL012346789</Value>
        </Identifier>
        <Documents>
          <Document>
            <Type>Factsheet</Type>
            <Date>2010-11-11</Date>
            <Language>NL</Language>
            <Url>http://example.org/factsheet.pdf</Url>
            <PublishedIn>
              <Country>NL</Country>
            </PublishedIn>
          </Document>
        </Documents>
      </Fund>
    </Funds>
  </Company>
</FundsDocumentXML>
```

1  
Figuur 5: XML structuur

FinFiles verwacht van de fondsaanbieder voor het invullen van het XML de volgende gegevens:

Element	Beschrijving
<b>CreationDate</b>	Datum van creatie van het XML bestand.
<b>&lt;Company&gt; Name</b>	Fondsaanbiedernaam.
<b>&lt;Fund&gt; Name</b>	Fondsnaam.
<b>&lt;Identifier&gt; Type</b>	Beschrijft het identificatietype van het fonds.
<b>&lt;Identifier&gt; Value</b>	Bevat de unieke code van een fonds.
<b>&lt;Document&gt; Type</b>	Kwalitatieve documentatie type: maandverslag, jaarverslag e.d.
<b>&lt;Document&gt; Date</b>	Bevat de publicatiedatum van het PDF document.
<b>&lt;Document&gt; Language</b>	Beschrijft de taal van het PDF document.
<b>&lt;Document&gt; Url</b>	Verwijzing naar het PDF document.
<b>&lt;PublishedIn&gt; Country</b>	Land van uitgifte.

Een fondsaanbieder (company) kan, zoals uit figuur 5 te lezen is, één of meer fondsen aanleveren. Elk fonds heeft zijn eigen unieke code, bijvoorbeeld de veel gebruikte ISIN-code<sup>2</sup>. Echter blijft uit de praktijk dat niet elk fonds een ISIN-code heeft. Daarom is er tijdens het opstellen van het XML formaat rekening gehouden met verschillende type identifiers.

Een fonds kan meerdere kwalitatieve documenten bevatten. Daarom zijn er per document een aantal elementen opgenomen, namelijk: type, date, language, url en de publishedIn country. De PDF documenten worden veelal in verschillende landen uitgegeven, tijdens het opstellen van het XML formaat is hier rekening mee gehouden.

Deze simpele maar doeltreffende XML biedt voldoende mogelijkheden om de kwalitatieve documenten gestructureerd op te slaan op een FinFiles server.

### 8.1.2 VALIDATIE AAN DE HAND VAN XML SCHEMA

Om het XML bestand te kunnen valideren, is er een XML schema opgesteld. Het XML schema legt restricties op diverse elementen: de types, languages en countries. Het schema zorgt ervoor dat fondsaanbieders zich houden aan het XML formaat, zodat er tijdens de verwerking van het XML bestand minder kansen op fouten ontstaan. Een fragment uit het XML schema wordt getoond in bijlage C.

<sup>2</sup> ISIN, International Securities Identifying Number, is een alfanumerieke code bestaande uit 12 karakter die diverse fondsen wereldwijd uniek identificeert.

## 8.2 VERWERKINGSENGINE

De Verwerkingsengine verwerkt de diverse cronjobs die zijn aangemaakt met het dashboard. De Verwerkingsengine bestaat uit twee verschillende acties, namelijk: GetFile en DownloadDocument. Dit hoofdstuk beschrijft de werking van deze twee acties.

### 8.2.1 GETFILE

Zodra een cronjob wordt verwerkt wordt er in eerste instantie gecontroleerd welke filetypes uit het XML bestand moeten worden behandeld. De parameters van de filetypes kunnen tijdens het aanmaken van de cronjobs worden opgegeven. Daarnaast kan aangegeven worden welke actie er moet worden uitgevoerd. Dit kan de GetFile of de DownloadDocument actie zijn. Hoe de actie GetFile werkt zal aan de hand van het volgende voorbeeld worden toegelicht:

Wanneer een cronjob wordt verwerkt worden de opgeslagen parameters uit de database uitgelezen. Deze parameters bevatten de acties die de cronjob uit gaat voeren en voor welke fondsaanbieder het geldt.

Wanneer de cronrecord de GetFile actie bevat gebeurt het volgende:

De tabel waar de XML locaties zijn opgeslagen wordt uitgelezen, waarbij de actieve locaties worden opgeslagen in een object. De verschillende XML locaties worden één voor één gevalideerd door de validateXML() methode (bijlage B1). Deze methode valideert het XML bestand aan de hand van het opgestelde XML schema.

Indien het XML bestand niet als valide wordt bestempeld worden de fouten gelogd. Met een valide XML bestand gaat het systeem verder met de verwerking. De logische opvolgende stap is het controleren van de locatie waar het XML bestand wordt aangeboden. Het systeem controleert het protocol van de verbinding. Het ontwikkelde systeem kan overweg met drie protocollen, namelijk http(s) en FTP.

In het geval van het http(s) protocol wordt de data uit het XML bestand opgeslagen in een object. Dit gebeurt met behulp van *simpleXML*. Wordt het FTP protocol gebruikt maakt het systeem, alvorens het opslaan van de gegevens, connectie met de FTP. Het dashboard biedt de mogelijkheid om tijdens op opslaan van de XML locatie een gebruikersnaam en wachtwoord mee te geven. De methode voor de FTP connectie wordt getoond in bijlage B1.

Zodra de fondsen met de bijbehorende gegevens uit het XML bestand is opgeslagen in een object, kan het systeem deze opslaan in de database. De gegevens worden opgeslagen in een tijdelijke tabel. Voor elk opgeslagen record <sup>3</sup> wordt er een verwerkingsstatus meegegeven, na het opslaan van de XML gegevens wordt deze status op 0 gezet. Dit betekent dat de verwijzingen die opgeslagen zijn in de tijdelijke tabel nog niet zijn behandeld. De verwerking van de verwijzingen gebeurt in de DownloadDocument actie. Zowel de fouten tijdens de validatie van het XML bestand als het opslaan van de bijbehorende gegevens worden tijdens de verwerking gelogd.

Indien de XML gegevens zijn verwerkt is de tijdelijke tabel gevuld met fondsen met de bijbehorende documentverwijzingen welke nog niet gedownload zijn.

---

<sup>3</sup> Een record representeert een fonds met de bijbehorende elementen uit het XML bestand.

### 8.3.2 DOWNLOADDOCUMENT

Wanneer de cronjob de actie DownloadDocument bevat worden de records uit de tijdelijke tabel verwerkt. Evenals bij de GetFile actie kan dit indien nodig per fondsaanbieder worden verwerkt.

DownloadDocument slaat de te verwerken records, met verwerkingsstatus 0, voor de geselecteerde aanbieder op in een object, waarna deze wordt verwerkt. De verwerking wordt gestart wanneer het systeem weet dat er een geldige verwijzing is opgeslagen. Om een verwijzing te controleren op geldigheid is de functie getHeaders() geschreven (bijlage B2).

Pas wanneer de verwijzing geldig is start de verwerking. In principe kan het systeem nu de content achter de verwijzing opslaan op de FinFiles server. Echter moet er nog rekening gehouden worden met het feit dat PDF documentatie bij meerdere fondsen kan passen. In de praktijk worden er veelal jaarverslagen gepubliceerd die voor diverse fondsen dienen. Het is de taak van het systeem, vorens het downloaden van de documenten, de server te controleren op het bestaan van het bestand. Daarom wordt voordat het bestand wordt opgeslagen gecontroleerd of deze uniek is. Van de content wordt een *MD5 hash* gegenereerd. Deze MD5 hash wordt gebruikt ter controle. Komt de MD5 hash nog niet voor in het systeem wordt het opgeslagen op de server. Het systeem maakt hiervoor automatisch de directory structuur (bijlage B2).

Om een geschiedenis van de kwalitatieve documentatie te creëren, wordt er tijdens het aanmaken van de directory structuur gekeken naar de datum van het PDF document. Onderstaand figuur (figuur 6) geeft een voorbeeld van de directory structuur, Waarbij elk element variabel is.



Figuur 6: Directory structuur

Mocht er bij het downloaden en opslaan van de documentatie fouten ontstaan worden deze door het systeem gelogd. In geval van slagen wordt de verwerkingsstatus op 1 gezet en de records uit de tijdelijke tabel opgeslagen in de definitieve tabel. De externe verwijzingen naar de documentatie worden in de definitieve tabel vervangen voor verwijzingen naar de PDF documenten op de FinFiles server.

De twee acties GetFile en DownloadDocument vormen samen de Verwerkingsengine. Door middel van de geïmplementeerde controles zal het systeem bestaan uit alleen correcte kwalitatieve PDF documentatie. Om de bruikbaarheid van de Verwerkingsengine te vergroten is er een bijbehorend dashboard ontwikkeld. Het hoofdstuk 'Ontwikkeling dashboard' gaat dieper in op het dashboard.

## 9. ONTWIKKELING DASHBOARD

Het dashboard bevat twee beheeronderdelen voor het beheren van cronjobs en XML locaties. Daarnaast zijn functionaliteiten die betrekking hebben op het loggen van de Verwerkingsengine geïmplementeerd. FinFiles heeft tijdens de realisatie van het project de afstudeerstudent vrij gelaten in de ontwikkeling van de functionaliteiten binnen het dashboard.

### 9.1 MANAGE CRONJOB













De manage cronjob functionaliteit geeft de datamanager de mogelijkheid om de cronjobs per fondsaanbieder te beheren. Het toevoegen van een cronjob gebeurt aan de hand van een formulier. Hier kan gemakkelijk de te runnen tijden worden opgegeven. Daarnaast kan aangegeven worden welke actie de cronjob moet bevatten, de GetFile of de DownloadDocument actie. Ook kunnen de te verwerken documenttypen worden geselecteerd. Hier moet teruggedacht worden aan het voorbeeld waarbij jaarverslagen jaarlijks moeten worden gecontroleerd en gedownload. Cronjobs kunnen daarnaast gemakkelijk worden gewijzigd en verwijderd.

Het voordeel van een cronjob is het instellen van periodieke acties. Voor het aanmaken van een cronjob kan de gebruiker de dag van de week aangeven, de dag van de maand, welke maanden deze actief moet zijn en op welk uur en minuut van de dag.

Door de cronjobs goed in te regelen wordt het systeem efficiënt gebruikt.

### 9.2 MANAGE XML LOCATION

Zodra fondsaanbieders FinFiles op de hoogte hebben gesteld van de beschikbaarheid van de XML locaties, moeten deze worden toegevoegd aan het systeem. Hiervoor is in het dashboard een beheerfunctionaliteit geïmplementeerd. Per fondsaanbieder kunnen XML locaties worden toegevoegd gewijzigd en verwijderd. Wanneer het XML bestand op een FTP wordt aangeboden, kan er een gebruikersnaam en wachtwoord worden toegevoerd. Zodat de Verwerkingsengine verbinding kan maken met de FTP pagina. Onderstaand screenshot geeft een beeld van de manage XML locations functionaliteit.

Manage Cronjobs	Manage XML Locations	Manage Documents	Logs					
Manage XML Locations	XML Locations → Kempen & Co							
Select dataprovider	URL	IsActive	Creation Date	Modified Date	XMLType	Username	Password	
Kempen & Co	http://abnc-lennert/VerwerkingsEngine/Test.xml	0	2010-11-19 16:45:45	2010-11-19 16:45:45	Documents	12312		  
BNP Paribas	http://abnc-lennert/VerwerkingsEngine/warner.xml	0	2010-11-15 15:47:12	2010-11-15 15:47:12	Documents			  
	e.g. http://finfiles.nl/finfiles.xml	0	2010-11-19 12:42:05	2010-11-19 12:42:05	Documents	password for ftp		  
	e.g. http://finfiles.nl/finfiles.xml	1	2010-11-19 13:51:54	2010-11-19 13:51:54	Documents	password for ftp		  
<div>Delete processed XMLsDelete processed failures XML</div>								

Figuur 7: Screenshot manage XML locations

### 9.3 DOCUMENTPROCESSINGLOG

Beschreven in het hoofdstuk “Verwerkingsengine” worden de gegevens uit het XML bestand opgeslagen in een tijdelijke tabel. De records die deze tabel bevat kunnen worden bekeken met de DocumentsProcessingLog functionaliteit. Hier kan op fondsaanbiederniveau de te verwerken records worden bekeken. In het overzicht kan gemakkelijk onderscheid worden gemaakt tussen de te verwerken record en gefaalde downloads. Figuur 7 bevat een screenshot van de documentprocessinglog.

Manage Cronjobs   Manage XML Locations   Manage Documents   Logs									
Manage Documents									
	Productname	Identifiser	Filetype	Language	Documentdate	Downloaded	Processdate	Message	MD5 Hash
1	Kempen Orange Fund N.V.	LU0084302339	Factsheet	NL	2010-10-01	✗		Not processed	
2	Kempen Orange Fund N.V.	LU0084302339	Semi annual report	NL	2010-10-01	✓	2010-11-12 11:45:09	OK	3ee9ea5b7d5cd56cfa82d50d1b2d613
3	Kempen Orange Fund N.V.	LU0084302339	Prospectus	NL	2010-10-01	✗		Not processed	
4	Kempen Orange Fund N.V.	LU0084302339	Factsheet	NL	2010-10-01	✗		Not processed	
5	Kempen Orange Fund N.V.	LU0084302339	Semi annual report	NL	2010-10-01	✓	2010-11-12 11:45:10	OK	3ee9ea5b7d5cd56cfa82d50d1b2d613
6	Kempen Orange Fund N.V.	LU0084302339	Prospectus	NL	2010-10-01	✗		Not processed	
7	Kempen Orange Fund N.V.	LU0084302339	Factsheet	NL	2010-10-01	✗		Not processed	
8	Kempen Orange Fund N.V.	LU0084302339	Semi annual report	NL	2010-10-01	✓	2010-11-12 09:20:38	OK	3ee9ea5b7d5cd56cfa82d50d1b2d613
9	Kempen Orange Fund N.V.	LU0084302339	Prospectus	NL	2010-10-01	✗		Not processed	
10	Kempen Orange Fund N.V.	LU0084302339	Factsheet	NL	2010-10-01	✗		Not processed	
11	Kempen Orange Fund N.V.	LU0084302339	Semi annual report	NL	2010-10-01	✓	2010-11-12 11:45:11	OK	3ee9ea5b7d5cd56cfa82d50d1b2d613
12	Kempen Orange Fund N.V.	LU0084302339	Prospectus	NL	2010-10-01	✗		Not processed	
13	Kempen Orange Fund N.V.	LU0084302339	Factsheet	NL	2010-10-01	✗		Not processed	
14	Kempen Orange Fund N.V.	LU0084302339	Semi annual report	NL	2010-10-01	✗		Not processed	

Figuur 8: Screenshot documentprocessinglog

### 9.4 XMLPROCESSINGLOG

Voor de verwerking van de XML locaties is er een aparte logfunctionaliteit in het dashboard opgenomen. Op aanbiederniveau kunnen de XML locaties worden bekeken. Er wordt bij de locaties getoond wanneer deze voor het laatst is verwerkt en kan er worden opgevraagd of deze verwerking is geslaagd. Is dit niet het geval wordt er getoond wat er mis is gegaan tijdens de verwerking.

Manage Cronjobs   Manage XML Locations   Manage Documents   Logs																													
XML ProcessLog		XML processStatus																											
<div>Select dataprovider</div> <div>Kempen &amp; Co</div> <div>BNP Paribas</div>		<table> <tr> <th></th><th>Dataprovider</th><th>XML LocationURL</th><th>ProcessDate</th><th>Succeeded</th></tr> <tr> <td>1</td><td>Kempen &amp; Co</td><td>http://abnc-lennert/VerwerkingsEngine/Test.xml</td><td>2010-11-10 13:45:26</td><td>✗</td></tr> <tr> <td>2</td><td>Kempen &amp; Co</td><td>http://abnc-lennert/VerwerkingsEngine/Test.xml</td><td>2010-11-10 14:45:57</td><td>✗</td></tr> <tr> <td>3</td><td>Kempen &amp; Co</td><td>http://abnc-lennert/VerwerkingsEngine/Test.xml</td><td>2010-11-10 14:46:21</td><td>✗</td></tr> <tr> <td>4</td><td>Kempen &amp; Co</td><td>http://abnc-lennert/VerwerkingsEngine/Test.xml</td><td>2010-11-10 13:47:44</td><td>✗</td></tr> </table>				Dataprovider	XML LocationURL	ProcessDate	Succeeded	1	Kempen & Co	http://abnc-lennert/VerwerkingsEngine/Test.xml	2010-11-10 13:45:26	✗	2	Kempen & Co	http://abnc-lennert/VerwerkingsEngine/Test.xml	2010-11-10 14:45:57	✗	3	Kempen & Co	http://abnc-lennert/VerwerkingsEngine/Test.xml	2010-11-10 14:46:21	✗	4	Kempen & Co	http://abnc-lennert/VerwerkingsEngine/Test.xml	2010-11-10 13:47:44	✗
	Dataprovider	XML LocationURL	ProcessDate	Succeeded																									
1	Kempen & Co	http://abnc-lennert/VerwerkingsEngine/Test.xml	2010-11-10 13:45:26	✗																									
2	Kempen & Co	http://abnc-lennert/VerwerkingsEngine/Test.xml	2010-11-10 14:45:57	✗																									
3	Kempen & Co	http://abnc-lennert/VerwerkingsEngine/Test.xml	2010-11-10 14:46:21	✗																									
4	Kempen & Co	http://abnc-lennert/VerwerkingsEngine/Test.xml	2010-11-10 13:47:44	✗																									
		<div>Delete processed XMLs</div> <div>Delete processed failures XML</div>																											

Figuur 9: Screenshot XMLprocessinglog

## 9.5 STATISTIEKEN

De statistieken pagina geeft de gebruiker een beeld van de data die de Fund Document Engine bevat. Hierbij kan gedacht worden aan een lijst met statistieken met de volgende waarden:

- Aantal fondsaanbieders;
- Overzicht fondsaanbieders met aantal fondsen;
- Aantal opgeslagen kwalitatieve PDF documenten;
- Aantal te verwerken externe verwijzingen;
- Aantal onjuiste XML locaties (XML bestanden);
- Verwerkingstijd per XML bestand;

## 10. TESTING

Voordat het systeem volledig in gebruik kan worden genomen, moet het systeem aan diverse tests worden onderworpen. De tests zorgen ervoor dat eventuele *bugs* vroegtijdig uit het systeem zijn.

Tijdens de ontwikkeling van de Verwerkingsengine zijn de verschillende onderdelen veelvuldig getest. Allereerst de aanlevering van de XML bestanden. Er zijn diverse XML bestanden lokaal en extern gehost, waarna er is gekeken of de content uit het XML bestand kon worden gevalideerd. De verschillende locaties maakte gebruik het http(s) of het FTP protocol. De resultaten van deze test waren positief.

Om de geschreven techniek achter de validatie te testen zijn er diverse XML bestanden, zowel foutief als correct, verwerkt door het systeem. Het systeem detecteerde de fouten waarna de fouten met een bijbehorende melding werden op geslagen in de database. Het verwachte slagingspercentage van 98% werd tijdens deze tests glansrijk gehaald.

De tests gericht op de Verwerkingsengine zijn tijdens de afstudeerperiode getest door de junior developer met ondersteuning van de afstudeerstudent. Het is voor FinFiles raadzaam om voordat het systeem live gaat een testopstelling met diverse fondsaanbieders te realiseren. Zoals beschreven gaat de ABN AMRO tool gebruik maken van het Open Fund Source. De testomgeving kan met de 13 fondsaanbieders, ontsloten in de ABN AMRO tool, worden opgezet.

Het dashboard is getest op usability. Hiervoor zijn er testpaden opgesteld. De resultaten van de testpaden voldeden aan de verwachte resultaten.

De Verwerkingsengine draaide tijdens de tests lokaal. Voor het testen van de performance van de Verwerkingsengine dient FinFiles het systeem te testen op de server waar het systeem op gaat draaien. Hierbij is het van belang dat de snelheid, waarmee de diverse PDF documenten worden gedownload, wordt bepaald. Deze snelheid is afhankelijk van de servers waar de fondsaanbieders de PDF documenten hosten. Het is belangrijk dat er getest wordt hoelang het duurt om diverse documenten te verwerken. Zodat de cronjobs effectief kunnen worden gemanaged, waardoor de server wordt ontlast.



## 11. IMPLEMENTATIE

Het implementeren van de Verwerkingsengine met het bijbehorende dashboard is tamelijk eenvoudig. De technische implementatie wordt in hoofdstuk 11.1 beschreven.

Het dashboard kan, zodra de technische implementatie is uitgevoerd, vanuit elke browser worden gebruikt. Gebruiksvriendelijkheid staat voor het gebruik van het dashboard voorop. Mocht de gebruiker onverwacht voor problemen komen te staan, kan de gebruiker de handleiding raadplegen.

### 11.1 TECHNISCHE IMPLEMENTATIE

Beschreven in de scriptie maakt de Fund Document Engine deel uit van het Open Fund Source. Na de oplevering van het project is echter het beheer van de kwantitatieve data niet volledig ontwikkeld. Daarom is er rekening mee gehouden dat de Fund Document Engine los in gebruik kan worden genomen. Het is hierbij belangrijk dat er geen mogelijkheid bestaat om fondsaanbieders toe te voegen aan het systeem, dit dient door FinFiles zelf te worden ontwikkeld.

Van de gebruikte server wordt verwacht dat deze beschikking heeft over:

- **Apache**  
Aanbevolen is de nieuwste versie van Apache te gebruiken, mits deze ondersteuning biedt voor de onderstaande versie van PHP. De applicatie is ontwikkeld op Apache 2.1.11.
- **PHP 5.3.0**  
De applicatie is ontwikkeld met PHP 5.3.0, na de installatie van PHP op de server moet een extra extensie worden geactiveerd.
  - cURL  
cURL is de vereiste PHP extensie. De extensie wordt gebruikt tijdens het opslaan van de PDF documentatie.
- **MySQL**  
De database zal moeten draaien op MySQL server met versie 5.1.36 of hoger om onverwachte problemen uit de weg te gaan.

De source van de Fund Document Engine wordt aangeleverd in een rar-bestand, welke op de web server dient te worden uitgepakt.

## 12. CONCLUSIE EN AANBEVELING

Met de opgeleverde Fund Document Engine is FinFiles klaar om het huidige Datasysteem te vervangen door het nieuwe geautomatiseerde systeem. Tijdens het project is er voldaan aan de eisen die FinFiles stelt aan het nieuwe systeem. Met de Fund Document Engine kan FinFiles betere kwaliteitscontrole uitvoeren op de kwalitatieve fondsdata. De Fund Document Engine is ontwikkeld om de kwalitatieve fondsdata door FinFiles te laten beheren. Het opslaan van de kwalitatieve fondsdata op eigen servers zorgt er voor dat een historie opgebouwd kan worden. FinFiles is op het gebied van datadistributie, afgezien van de aanlevering van XML bestanden, niet meer afhankelijk van welke data de fondshuizen op dat moment aanbieden.

Met behulp van het beheersysteem voor de cronjobs en XML locaties kan FinFiles gemakkelijk de Verwerkingsengine bedienen, de dashboardfunctionaliteiten geven op hun beurt inzicht in de processen die door de Verwerkingsengine worden uitgevoerd.

FinFiles streeft naar een succesvolle verwerking van 98% van de geleverde kwalitatieve fondsdata. De diverse tests die uitgevoerd zijn, wijzen uit dat de 98% haalbaarheid geen utopie is. Fouten tijdens de verwerking worden gelogd, de logs kan FinFiles in de toekomst uitbreiden met extra functionaliteiten. Hierbij kan gedacht worden aan het automatisch versturen van de logs naar de fondsaanbieder zodat zij het XML bestand kunnen aanpassen. Het toevoegen van kleine functionaliteiten zoals het versturen van automatische e-mailberichten komen het systeem aanzienlijk ten goede.

De verschillende document types worden op het moment via *PHPmyAdmin* toegevoegd aan de database. Voor FinFiles is het een kleine moeite om een simpele beheerfunctionaliteit binnen het dashboard te implementeren waarmee de document types gemakkelijk beheert kunnen worden.

### 13. EVALUATIE

Terugkijkend op de gehele afstudeerperiode kan ik vol enthousiasme zeggen dat ik veel nieuwe kennis heb opgedaan. Zoals al te lezen was in het voorwoord, heb ik tijdens de opleiding enkele keren last gehad van motivatie problemen. Wat leidde tot het niet afronden van de belangrijke vakken. Hierdoor kwam mijn kennis met betrekking tot programmeren achter te liggen op wat er verwacht wordt van de opleiding. Mede daardoor vond ik de opleiding minder aantrekkelijk worden en kreeg ik het gevoel dat Informatica voor mij niet de juiste opleiding was. Tijdens de afstudeerperiode is het tegendeel bewezen.

Gelijktijdig met het schrijven van het onderzoeksrapport, functioneel en technisch ontwerp heb ik mijzelf meer proberen te ontwikkelen in het programmeren. Dit heb ik gedaan om tijdens de realisatiefase niet voor grote problemen te komen staan. Tijdens het programmeren heb ik mij bezig gehouden met diverse functionaliteiten die het afstudeerproject zou gaan bevatten. Mijn afstudeerbegeleider heeft mij hier goed in gestuurd waardoor ik de te gebruiken programmeertaal PHP beter onder de knie kreeg. De feedback van mijn begeleider resulteerde veelal in het herschrijven van de desbetreffende code waardoor ik al voor de realisatie goede inzichten kreeg in het te ontwikkelen product. Deze activiteiten zorgden ervoor dat ik met plezier aan het programmeren was, wat in voorgaande jaren niet altijd het geval is geweest.

Het plezier en de motivatie hebben uiteindelijk deze scriptie en een goed werkend product als resultaat gebracht. Tijdens de afstudeeropdracht heb ik PHP programmeren goed onder de knie gekregen, daarnaast heb ik geleerd om diverse API's te gebruiken bij het programmeren. Ook het toepassen van een goed MVC-model en het gebruiken van een framework gaan mij goed af.

Naast de grote stappen die ik gemaakt heb met betrekking tot programmeren, heb ik tijdens de afstudeerperiode veel geleerd van het bedrijfsleven. Binnen FinFiles ben ik betrokken bij verschillende lopende projecten. Werkvergaderingen werden door mij bijgewoond en naast mijn afstudeeropdracht kreeg ik diverse andere verantwoordelijkheden.

De goede begeleiding vanuit FinFiles heeft ervoor gezorgd dat dit resultaat is neergezet. In geval van grote knelpunten werd ik bijgestaan door mijn afstudeerbegeleider of de junior developer. Zij gaven mij tips waardoor ik de knelpunten zelf kon oplossen. De afstudeerperiode is voor mij een uitdagende en leerzame periode geweest waar uiteindelijk een waardevol product is uitgekomen.

## 14. VERKLARENDE WOORDENLIJST

Term	Verklaring
<b>Asset Manager</b>	Vermogensbeheerder, voor het beheer van groot vermogens voor derden door gespecialiseerde organisaties of personen.
<b>Cronjobs</b>	Commando's om op specifieke tijdstippen scripts te laten runnen. Veelal gebruikt binnen Unix-achtige systemen.
<b>CSS</b>	Cascading Style Sheet, wordt gebruikt voor de vormgeving van een HTML bestand.
<b>Database</b>	Digitaal archief voor het opslaan van uiteenlopende gegevens.
<b>Datasysteem</b>	Het Datasysteem is het huidige beheersysteem welke FinFiles gebruikt voor het managen van de kwalitatieve data.
<b>E-Business</b>	Kort gezegd bevat E-Business alle zaken met betrekking tot het online zaken doen, waarbij bedrijven, consumenten en de overheid bij betrokken zijn.
<b>Fondsaanbieder</b>	Zie Asset Manager.
<b>Framework</b>	Een framework is een geheel van softwarecomponenten dat gebruikt kan worden bij het programmeren van applicaties.
<b>Fund Data Interface</b>	Een product van FinFiles waarmee op een gebruiksvriendelijke manier fondsinformatie kan worden opgevraagd en vergeleken.
<b>Fund Document Engine</b>	Benaming voor het project wat de student tijdens zijn afstudeerperiode heeft gerealiseerd.
<b>HTML</b>	Hyper Text Markup Language, HTML is een opmaaktaal voor de specificatie van gegevens.
<b>IDE</b>	Integrated Development Environment, een IDE is ontwikkelingsomgeving voor developers.
<b>Kwalitatieve data</b>	Publicaties over fondsen, hierbij moet gedacht worden aan maandverslagen, jaarverslagen, prospectussen en dergelijke.
<b>Kwantitatieve data</b>	Dagelijkse fondsgegevens, hierbij moet gedacht worden aan NAVs, rendementen en bijvoorbeeld risico's.
<b>Loggingdashboard</b>	Dashboard/beheersysteem, waarmee de Verwerkingsengine kan worden beheert en gemonitord.
<b>MD5 Hash</b>	MD5 is een hashfunctie die wordt gebruikt om de integriteit van bestanden te controleren.
<b>MySQL</b>	MYSQL is een Open Source relationele databasemanagementsysteem dat gebruikt van SQL.
<b>Open Fund Source</b>	Benaming voor de innovatieve applicatie, waarbij kwalitatieve en kwantitatieve

	data op één plek wordt gemanaged.
<b>Open Source</b>	Open Source beschrijft de praktijk die in productie en ontwikkeling vrije toegang geeft tot de bronmaterialen van het eindproduct.
<b>PHP</b>	PHP: Hypertext Preprocessor. PHP is een scriptaal die bedoeld is om dynamische webpagina's te creëren.
<b>PHPMyAdmin</b>	PHP programma om My-SQL databases te beheren.
<b>SQL</b>	SQL is een gestandaardiseerde taal waarmee aanvragen op een relationele database kunnen worden uitgevoerd.
<b>Usability</b>	Gebruiksvriendelijkheid van een softwareprogramma, waarbij gedacht kan worden aan logische workflows.
<b>Use case diagram</b>	Een use case diagram beschrijft de functionaliteiten die binnen een applicatie worden gerealiseerd.
<b>Verwerkingsengine</b>	Onderdeel van de Fund Document Engine, zorgt voor de verwerking van de XML bestanden.
<b>XML</b>	Extensible Markup Language, een standaard om gestructureerde gegevens te weergeven in vorm van platte tekst.
<b>XML schema</b>	Een XSD, XML Scheme Definition, is een taal voor het beschrijven van de structuur van XML documenten.

## 15. BRONVERMELDING

### 15.1 INTERNETBRONNEN

Eclipse, *Eclipse for PHP developers*, geraadpleegd op 1 september 2010,  
<http://www.eclipse.org/downloads>

WAMP, *WampServer 2.0i*, geraadpleegd op 1 september 2010,  
<http://www.wampserver.com/en/>

w3schools, *XML Schema Tutorial*, geraadpleegd op 15 september 2010,  
<http://www.w3schools.com/schema/default.asp>

PHP.net, *Document Object Model*, geraadpleegd op 27 september 2010,  
<http://php.net/manual/en/book.dom.php>

PHP.net, *SimpleXML*, geraadpleegd op 27 september 2010,  
<http://nl2.php.net/manual/en/book.simplexml.php>

PHP.net, *FTP*, geraadpleegd op 8 oktober 2010,  
<http://nl.php.net/manual/en/book.ftp.php>

PHP.net, *Client URL Library*, geraadpleegd op 15 oktober 2010,  
<http://nl.php.net/manual/en/book.curl.php>.

Smarty, *Smarty template engine*, geraadpleegd op 1 november 2010,  
<http://www.smarty.net/documentation>

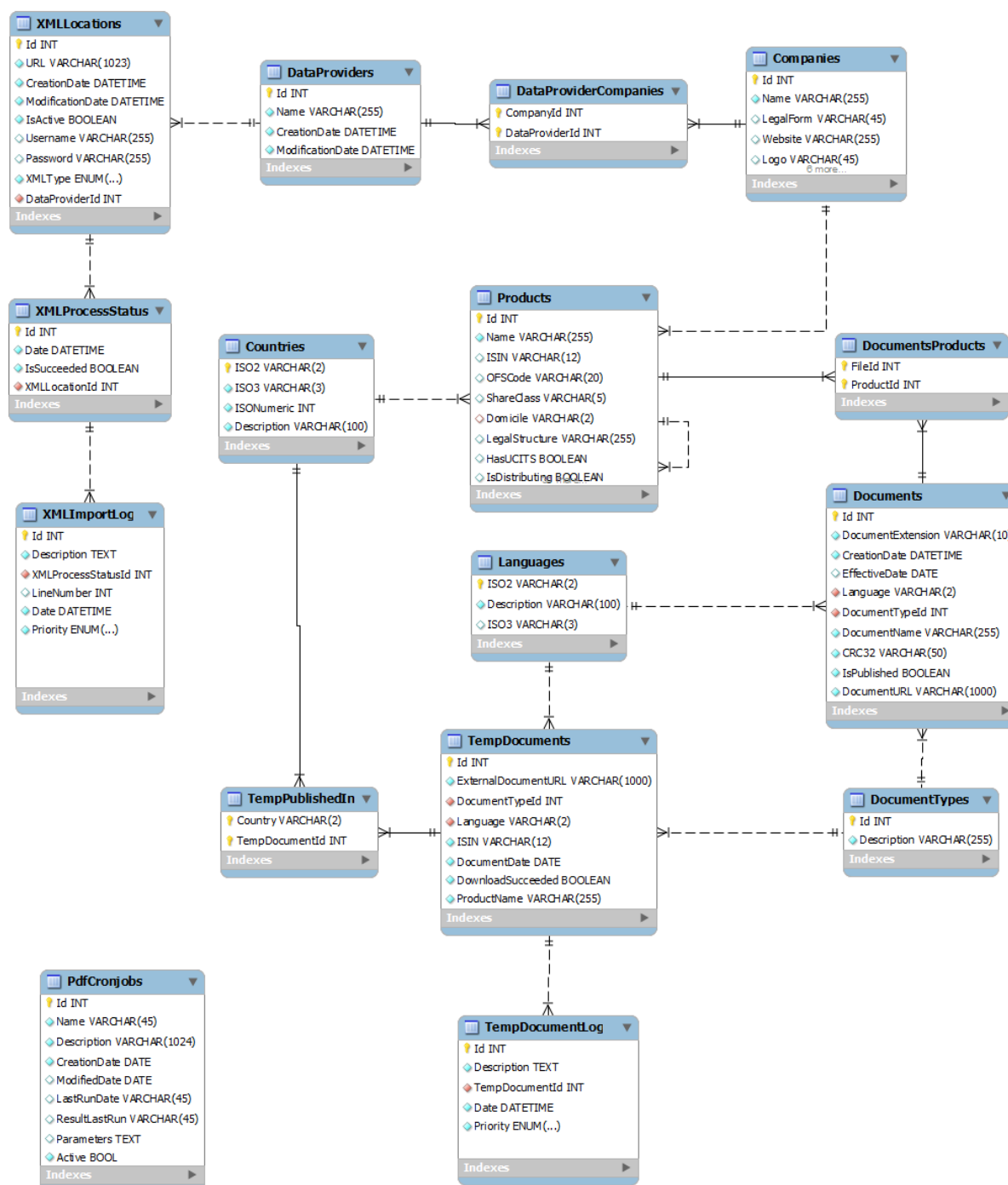
### 15.2 LITERATUUR

Leren Communiceren druk 5, *Michaël Steehouder Noordhoff Uitgevers B.V.*,  
geraadpleegd op 22 november 2010

## 16 BIJLAGEN

## BIJLAGE A: DATABASEMODEL

Onderstaand database model maakt deel uit van het databasemodel voor het Open Fund Source. Het model geeft een overzicht van de gebruikte tabellen voor de Verwerkingsengine en het dashboard.



## BIJLAGE B: GEBRUIKTE METHODEN

De belangrijkste gebruikte methoden voor de ontwikkeling van de Verwerkingsengine worden in bijlage B weergegeven.

### BIJLAGE B1: GETFILE

#### XML validatie functie

De functie voor het valideren van de XML bestanden controleert het bestand achter de XML locatie. Dit gebeurt aan de hand van het XML schema (bijlage C). Eventuele foutmeldingen worden na de validatie opgeslagen in de database.

```
public function validateXML($url, $xsd) {
    libxml_use_internal_errors ( true );
    $objDom = new DOMDocument ( );
    $objDom->load ( $url );
    if ( ! $objDom->schemaValidate ( $xsd ) ) {
        $this->_errors = libxml_get_errors ( );
        new Debug($this->_errors);
        return false;
    }
    return true;
}
```

#### Connectie FTP protocol

Onderstaande functie maakt met behulp van de FTP gegevens en de XML locatie verbinding met de FTP. Waarna na het opzetten van een succesvolle verbinding de gegevens uit de XML worden verwerkt.

```
public function FtpConnect($ftpUser, $ftpPass, $url) {
    $this->_url = $url;
    // Parse_url for getting the host
    $ftpServer = parse_url ( $url );
    $this->_connectionId = ftp_connect ( $ftpServer ['host'] );
    $loginResult = ftp_login ( $this->_connectionId, $ftpUser,
    $ftpPass );
    ftp_pasv ( $this->_connectionId, true );
    if ( (! $this->_connectionId) || (! $loginResult) ) {
        return false;
    } else
        return true;
}
```



**BIJLAGE B2: DOWNLOADDOCUMENT****getHeaders**

In enkele gevallen verwijzen de URLs naar de PDF documenten door naar een andere link. Om het correcte bestand te downloaden is de functie `getHeaders` geschreven.

```
function getHeaders($externalDocumentURL) {
    //Do some formatting
    $externalDocumentURL = str_replace ( ' ', '%20',
    $externalDocumentURL);

    $handle = curl_init ( $externalDocumentURL );
    curl_setopt ( $handle, CURLOPT_RETURNTRANSFER, TRUE );
    curl_setopt ( $handle, CURLOPT_FOLLOWLOCATION, TRUE );
    $response = curl_exec ( $handle );
    $header = curl_getinfo ( $handle );
    return $header;
    curl_close ( $handle );
}
```

**createDirectory**

Wanneer aan het Open Fund Source nieuwe aanbieders, fondsen en bestanden worden toegevoegd. Moet er een directory structuur worden aangemaakt. `CreateDirectory` creëert een directory indien deze nog niet bestaat.

```
function createDirectory($dir, $mode = 0755,$recursive=true) {
    if(mkdir($dir,$mode,$recursive))
        return true;
}
```

## BIJLAGE C: FRAGMENT FUNDDOCUMENT.XSD

Onderstaand fragment maakt deel uit van het opgestelde XML schema. Het XML schema wordt gebruikt ter validatie van de aangeleverde XML bestanden.

```
<xs:element name="Documents">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="Document" maxOccurs="unbounded">
        <xs:complexType>
          <xs:sequence>
            <xs:element name="Type" type="xs:string">
              <xs:annotation>
                <xs:documentation xml:lang="en">
                  Specifies the documenttype e.g., factsheet,
                  annual report, interim report, financiële bijsluiter.
                </xs:documentation>
              </xs:annotation>
            </xs:element>
            <xs:element name="Date" type="xs:date">
              <xs:annotation>
                <xs:documentation xml:lang="en">
                  Specifies the creation date of the document
                </xs:documentation>
              </xs:annotation>
            </xs:element>
            <xs:element name="Language">
              <xs:annotation>
                <xs:documentation xml:lang="en">
                  Specifies the language of the document language
                  (ISO 639-1 alpha-2)
                </xs:documentation>
              </xs:annotation>
              <xs:simpleType>
                <xs:restriction base="xs:string">
                  <xs:length value="2" />
                </xs:restriction>
              </xs:simpleType>
            </xs:element>
            <xs:element name="Url" type="xs:string">
              <xs:annotation>
                <xs:documentation xml:lang="en">
                  The URL to the location of the document e.g.,
                  http://example.org/documents/factsheet_fundsA_dec2010.pdf
                </xs:documentation>
              </xs:annotation>
            </xs:element>
            <xs:element name="PublishedIn">
              <xs:complexType>
                <xs:sequence>
                  <xs:element name="Country" maxOccurs="unbounded">
                    <xs:annotation>
                      <xs:documentation xml:lang="en">
                        Specifies the countries in which the document
                        may be published. (ISO 3166-1 alpha-2)
                      </xs:documentation>
                    </xs:annotation>
                    .....
```

## BIJLAGE D: SOURCE STRUCTUUR

Onderstaand figuur geeft de structuur van de Verwerkingsengine weer. De structuur houdt zich aan het MVC-model.

