



Hasso-Plattner-Institut für IT-Systems Engineering

Bachelorarbeit

Eine Architektur für ein ereignisgesteuertes, webbasiertes Backend für Project-Zoom

An eventdriven webbased backend architecture for Project-Zoom

Tom Bocklisch

tom.bocklisch@student.hpi.uni-potsdam.de

Betreut von Prof. Dr. Holger Giese,
Thomas Beyhl, M.Sc. und
Gregor Berg, M.Sc.

FG Systemanalyse und Modellierung

Potsdam, 01. July 2013

Zusammenfassung

Im Verlauf der kreativen Design-Prozesse in der HPI School of Design Thinking (D-School) entstehen eine Menge an DokumentationsArtefakten. Diese Daten sind jedoch meist in Quellen abgelegt, deren gesicherte Persistenz nur für einen beschränkten Zeitraum gegeben ist. Die Informationen aus den Wissensquellen gilt es dauerhaft zu sichern, zu organisieren und für Außenstehende verständlich zu präsentieren. Durch die graphische Darstellung des Prozesses der Ideenfindung können sowohl die Teilnehmer eines Design Thinking-Teams, als auch deren Mentoren, wichtige Erkenntnisse für den weiteren Erfolg einer Idee gewinnen. Neben diesem Überblick über die Entwicklungen in einem Projekt, fehlt der D-School eine Gesamtübersicht über alle Projekte. Diese hilft, Abhängigkeiten der Projekte voneinander zu verdeutlichen, und erleichtert die Akquirierung neuer Projekt-Themen.

In dieser Bachelorarbeit soll das Backend der entwickelten Lösung *Project-Zoom* näher erläutert werden. Dabei werden die Architektur des Systems betrachtet und die in diesem Zusammenhang zugrundeliegenden Entscheidungen erläutert. Dazu werden zu Beginn der Arbeit die Anforderungen des Projektpartners aufgezeigt, welche im Verlauf des Entwicklungsprozesses als Entscheidungsgrundlage galten. Das Backend der umgesetzten Anwendung basiert auf einer asynchronen Webarchitektur. Es wird das Konzept des data-centric Designs umgesetzt und mit verschiedenen alternativen Konzepten im Bereich der Datenbankanbindung verglichen. Die Datenmodellierung setzt die Anforderungen der HPI School of Design Thinking um und verwendet dabei unter anderem Datenversionierung und Datenzugriffsschutz auf Datenmodell-Ebene. Zur Anbindung von externen Komponenten und Systemen wurde ein Eventsystem umgesetzt, welches die Anpassung der Anwendung an zukünftige Bedürfnisse der School of D-School erlaubt. Den Abschluss der Arbeit bilden eine Evaluierung der Anforderungen sowie ein Ausblick auf mögliche Erweiterungen von Project-Zoom.

Inhaltsverzeichnis

1	Motivation	1
1.1	Systemanforderungen	2
1.2	Idee von Project-Zoom	4
1.3	Abgrenzung	5
2	Architektur des Backends	7
2.1	Anwendungsart	7
2.2	Datenmodellierung	9
2.3	Data-centric Design	11
2.4	Eventsystem	18
3	Umsetzung des Backends	21
3.1	Technologieauswahl	21
3.2	Data-centric Design in Project-Zoom	26
3.3	Anbindung externer Komponenten	27
3.4	Evaluierung der Anforderungen	29
4	Fazit und Ausblick	31
	Literaturverzeichnis	33
	Glossar	37
A	Quelltexte	A-1
A.1	Quelltext-Beispiel für eine JSON-Transformation und -Validierung	A-1
B	Interview Protokolle	B-1
B.1	Interview 22.11.2012 - Claudia Nikolai	B-1
C	Ergänzende Visualisierungen	C-1
C.1	Datenmodell von Project-Zoom	C-1
C.2	Objektdiagramm für einen beispielhaften Graphen	C-2
C.3	InfoQ Umfrage zu Webframeworks der JVM	C-2
C.4	Performance Test	C-3
C.5	Überblick über die vorhandenen Events	C-5
D	REST-Schnittstelle	D-1

1 Motivation

Die HPI School of Design Thinking lehrt die kreative Herangehensweise an Probleme und die Entwicklung einfallsreicher Lösungen. Die Lehre findet dabei in verschiedenen Kursen statt. Studierende können sich die Grundlagen im Basic-Track aneignen und ihr Wissen nach erfolgreicher Absolvierung im Advanced-Track vertiefen. Neben den studentischen Kursen bietet die D-School auch Weiterbildungskurse für Unternehmen an.

In der Regel besteht ein Design Thinking-Kurs aus mehreren multidisziplinären Kleingruppen von üblicherweise sechs Teilnehmern, die gemeinsam an einem Projekt arbeiten. Das Projektthema wird dabei entweder von der D-School selbst oder, vor allem bei längeren Projekten, von externen Projektpartnern vorgeschlagen.

Bei der Arbeit an einem Projekt durchlaufen die Teilnehmenden verschiedene Phasen, die in der Abbildung 1.1 dargestellt sind: Während zunächst das Verstehen und Beobachten des Problemfeldes im Vordergrund stehen, wird anschließend ein Standpunkt definiert. Es folgt die Phase der Ideenfindung, an die die Erstellung eines Prototypen anknüpft. Dieser Prototyp wird abschließend mit geeigneten Versuchspersonen getestet (vgl. [PMW09]). Dabei entstehen im Verlauf der Phasen unterschiedliche Dokumente, welche die Lösungsfindung dokumentieren.

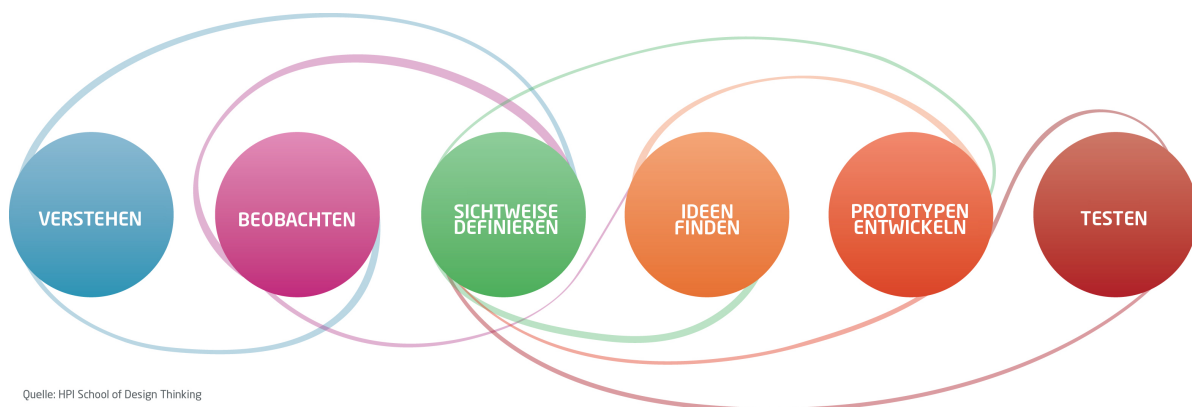


Abbildung 1.1 – Phasen des D-School-Prozesses¹

Im Verlauf des Projektes kommt es durchaus vor, dass ein Team eine Phase mehrmals durchläuft oder in eine vorherige Phase zurückkehrt. Dies erschwert die Organisation der Dokumente, z.B. in einer einfachen hierarchischen Struktur. Weiterhin ist es schwierig, allein aus den Dokumenten deren Entstehungsreihenfolge und Bedeutung zu erfassen. Meist entstehen während eines

¹Quelle [PMW09, p. 114]

drei Monate andauernden Projektes verschiedenste Präsentationen, Zusammenfassungen, Prototypen, Bilder von Whiteboards und Interviewdokumentationen. Diese werden in der Regel in der von den Studierenden bevorzugten Art und Weise gespeichert und verwaltet, beispielsweise mit Hilfe von *Dropbox*², *Google Docs*³ oder *Box*⁴.

Das Verständnis der Dokumentation ist sowohl für das Projekt selbst, zum Verstehen und Erlernen des Prozesses, als auch als Ideenquelle für zukünftige Projekte wichtig. Ferner sind die Projektergebnisse nützlich, um neue Projektpartner anzuwerben, welche zukünftige Projekte anbieten.

1.1 Systemanforderungen

Im Verlauf des Projektes wurden in Zusammenarbeit mit den Mitarbeitern und Studierenden der D-School verschiedene Anforderungen an Project-Zoom als Projektverwaltungs- und Dokumentationstool aufgestellt (vgl. [BBD⁺13]). Aus diesen leiten sich die umgesetzte Architektur und die verwendeten Technologien ab. In dieser Arbeit aufgeführt sind nur die für diesen Teil des Projektes wichtigen und relevanten Anforderungen. Für eine Analyse der Nutzergruppen und deren Bedürfnisse sei auf [BBD⁺13] verwiesen.

Funktionale Anforderungen

F1: Es muss sichergestellt werden, dass kein unautorisierter Nutzer auf Daten Zugriff hat, die geschützt sind.

Dies ergibt sich aus dem Vorhandensein von Geheimhaltungsverträgen (NDA) für Projekte und dem Wunsch der D-School den Zugriff auf Projekte einzuschränken.

F2: Innerhalb der Anwendung sollen die Daten aus *Filemaker*⁵, *Box*, Netzwerkdateisystemen⁶ und *Facebook*⁷ aggregiert werden.

Die angegebene Reihenfolge entspricht dabei dem Wunsch der Umsetzung von der D-School.

²Dropbox, <http://www.dropbox.com> (Zugriff 13.06.13)

³Google Docs, <http://drive.google.com> (Zugriff 12.06.13)

⁴Box, <http://box.com> (Zugriff 13.06.13)

⁵Filemaker ist die Datenbank, welche die D-School verwendet um Projekt und Userdaten zu speichern. <http://www.filemaker.com> (Zugriff 29.06.13)

⁶Ein Netzwerkdateisystem ist ein Dateisystem, auf welches von Rechnern im selben Netzwerk zugegriffen werden kann, um zum Beispiel Dateien zu lesen.

⁷The Facebook, <http://www.facebook.com> (Zugriff 29.06.13)

F3: Aggregierte Daten sollen nicht gelöscht werden.

Es kann durchaus vorkommen, dass Daten in aggregierten Quellen nach einiger Zeit gelöscht werden müssen. Die Anwendung soll die Daten dann weiterhin vorhalten können. Dass eine Ressource in der originalen Quelle nicht mehr vorhanden ist, kann zum Beispiel durch eine Markierung in einer Datenbank erfolgen.

F4: Die Anbindung von neuen Datenquellen muss möglich sein. Das Core-Backend soll dabei unverändert bleiben.

Über die Zeit wandeln sich die von den Studierenden verwendeten Werkzeuge zur Dokumentation. Dies muss bei der Anwendungskonzeption berücksichtigt werden.

F5: Es dürfen keine aggregierten Daten in deren Quellen verändert werden.

Die Anwendung soll auf externe Daten nur lesend zugreifen. Dies dient dem Schutz der Datenquellen.

F6: Die Studierenden sollen in der Lage sein, die Anwendung außerhalb der D-School verwenden zu können.

Dies ist wichtig, da die Studierenden meist nur zwei Tage in der Woche in der D-School verbringen. Zum Teil treffen sich die Projektteilnehmer außerhalb der D-School mit Interview- und Projektpartnern, welche am Stand des Projektes interessiert sind.

Nichtfunktionale Anforderungen

NF1: Es sollen 100 Nutzer gleichzeitig in der Lage sein, das Backend zu nutzen.

Die Anzahl von 100 Nutzern ergibt sich als Summe der maximale Anzahl aller Teilnehmenden und Lehrenden. Da keine anderen Nutzer Zugriff auf das System haben sollen, entspricht dies einer oberen Schranke für die Anzahl der aktiven Nutzer.

NF2: Der Server muss die Updates aus den Datenquellen innerhalb von einer Minute verarbeiten.

Umso schneller die Updates vom System verarbeitet werden, desto eher können die Nutzer mit den Daten in der Anwendung arbeiten. Die Grenze von einer Minute wurde gewählt, da dies einerseits eine realistische Zeitspanne für notwendige Berechnungen ist und andererseits den Studierenden das reibungslose Arbeiten mit dem System ermöglicht.

1.2 Idee von Project-Zoom

Im Entwicklungsprozess einer Lösung zur Verbesserung der Dokumentation in der D-School wurde auf den D-School-Prozess zurückgegriffen. Diese Vorgehensweise ermöglichte einen sehr guten Einblick in die Arbeitsweise der Teilnehmer in den Projekten der D-School.

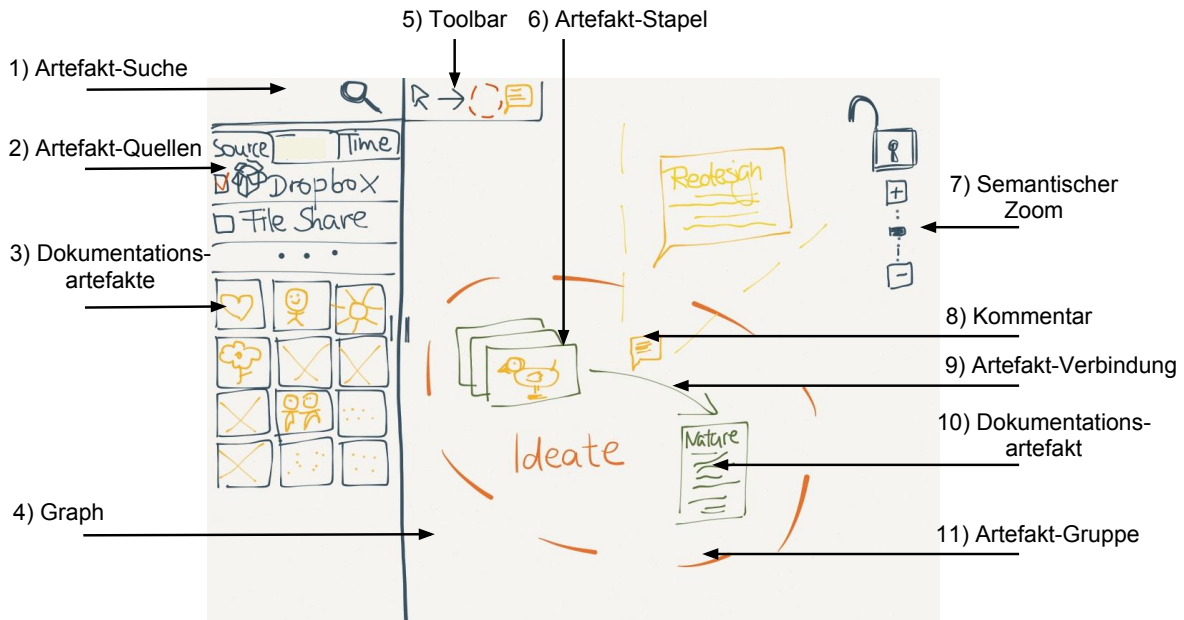


Abbildung 1.2 – Prototyp einer Übersicht über den Dokumentationsprozess eines Projektes

In mehreren Iterationen wurde Project-Zoom entwickelt und verfeinert. Einer der ersten Entwürfe der Nutzoberfläche ist in Abbildung 1.2 zu sehen. Das Konzept von Project-Zoom ist die Verbindung einer Projektübersicht mit einem Projekteinblick. Der Nutzer soll die Möglichkeit haben einen Überblick über alle Projekte, und bei Bedarf einen tieferen Einblick in ein konkretes Projekt, zu bekommen.

Project-Zoom soll der Verbesserung der Dokumentation dienen. Dazu sollen die von den Studierenden erzeugten Artefakte durch manuelles Anordnen in eine Form gebracht werden, welche den Prozess der Gruppe visualisiert. Die Mitarbeiter der D-School können die entstandenen Graphen anschließend nutzen, um die Projektverläufe zu analysieren und gegebenenfalls Abläufe innerhalb einer Design Thinking-Phase anzupassen (vgl. Interview mit Claudia Nikolai im Anhang B.1).

Für eine reibungslose Integration des Systems ist vor allem die Anbindung an bereits existierende IT-Systeme und die von den Studierenden für das Projekt verwendete Software wichtig (vgl. Anforderung F2). Änderungen an diesen externen Diensten und der Auswahl der anzubindenden Datenspeicher ist über die Zeit sehr wahrscheinlich. Deshalb muss das System modular

aufgebaut sein, um nachträglich neue Dienste anbinden zu können, ohne den Backend-Core zu verändern.

1.3 Abgrenzung

Die Arbeit beschreibt und bezieht sich auf das Bachelorprojekt „From Creative Ideas to Well-Founded Engineering“ und das umgesetzte Softwaresystem Project-Zoom. Insgesamt haben sechs Studierende an dem Projekt gearbeitet, die in ihren Bachelorarbeiten Project-Zoom aus verschiedenen Blickwinkeln und mit verschiedenen Schwerpunkten beschreiben.

Tom Herolds Arbeit [Her13] beinhaltet die Interaktion mit kontextsensitiven Graphen. Dabei geht es darum, den Umgang der Studierenden mit der Nutzeroberfläche so intuitiv wie möglich zu gestalten und den Nutzer bei der Erfassung dokumentationsrelevanter Eigenschaften zu unterstützen.

Die Ausführungen von Anita Diekhoff [Die13] beschäftigen sich mit der Übersicht über Projekte und gehen näher auf Layoutfunktionalitäten in interaktiven Graphen ein.

Norman Rzepkas Bachelorarbeit [Rze13] thematisiert die webbasierte, eventgesteuerte, client-seitige Architektur von Project-Zoom. Dabei wird näher darauf eingegangen, wie die Daten von der Datenbank über das Backend asynchron an den Client ausgeliefert und angezeigt werden.

Die Arbeit von Dominic Bräunlein [Brä13] erläutert das Generieren und Bereitstellen von semantischen Thumbnails, um dem Nutzer das Erkennen der Dokumente seines Projektes zu erleichtern und somit selbst bei wenig verfügbarem Platz so viele Informationen eines Dokumentes anzeigen zu können wie möglich.

Thomas Werkmeisters Arbeit [Wer13] befasst sich mit der Anbindung externer Systeme zur Integration von Daten. Diese aggregierte Datenbasis ist die Grundlage für die Wissensbasis und die einzelnen Projekte.

Die Abbildung 1.3 zeigt einen groben Überblick über die Architektur des Systems. Der für diese Arbeit relevante Teil ist dabei durch einen gestrichelten Rahmen gekennzeichnet. Die Komponenten, die in den Arbeiten [Wer13] (Konnektoren) und [Brä13] (Thumbnailgenerator) beschrieben sind, sind mit dem hier erläuterten Systemteil mittels eines Eventsystems verbunden. Das Client-Frontend ist über eine REST-Anbindung⁸ an das Server-Backend angeschlossen. Mit der clientseitigen Implementierung der REST-Schnittstelle beschäftigt sich [Rze13].

In dieser Arbeit wird zunächst ein Überblick über das Gesamtsystem gegeben. Dazu werden die Anforderungen der D-School an das Backend analysiert, welche als Grundlage für die Wahl

⁸vgl. Konzept des Representational State Transfer in [Fie00]

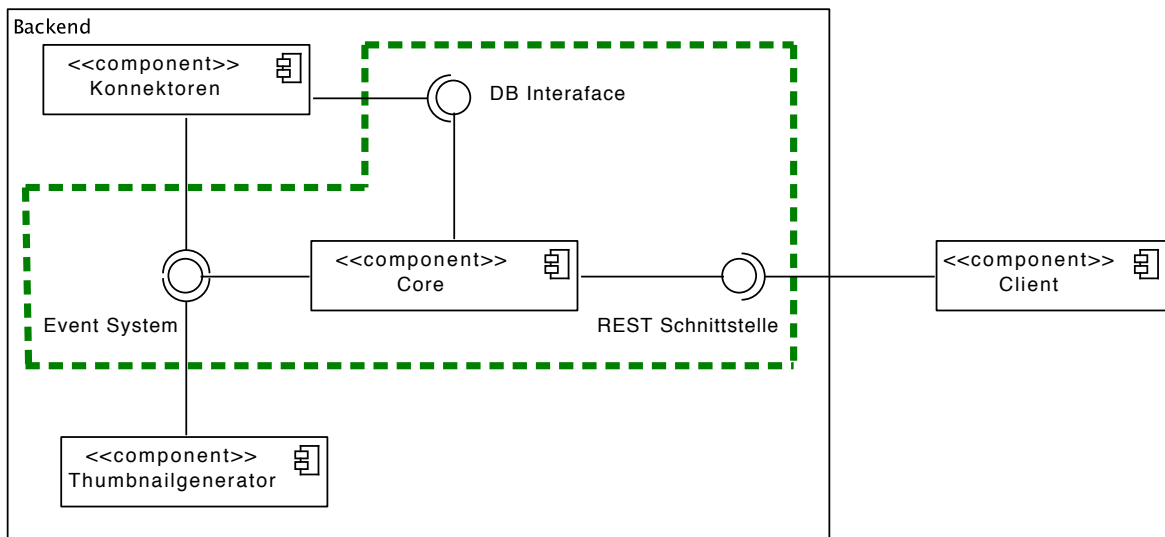


Abbildung 1.3 – Überblick über die verschiedenen Pakete und ihre Schnittstellen; in dem gestrichelten Rahmen der für diese Arbeit relevante Teil der Architektur

der verwendeten Architektur dienen. Anschließend wird die Architektur des Backends näher erläutert. Hier liegt der Hauptfokus zunächst auf einer neuen Art und Weise, eine Datenbank an eine Webapplikation anzubinden. Im Anschluss werden einige Feinheiten und ausgewählte Stellen der Umsetzung von Project-Zoom beleuchtet. Zum Abschluss werden die aufgestellten Anforderungen auf ihre Umsetzung in Project-Zoom geprüft.

2 Architektur des Backends

Inhalt dieses Kapitels ist die Betrachtung des Backend-Konzeptes von Project-Zoom. Begonnen wird mit einem Überblick über den Anwendungsaufbau. Anschließend wird ein konkretes Datenmodell entworfen, um daraufhin mit dem data-centric Design ein Konzept zur Umsetzung dieses Modells zu verdeutlichen. Den Abschluss bildet ein weiterer Teil der Architektur, das Eventsystem.

2.1 Anwendungsart

2.1.1 Webanwendung vs. native Applikation

Eine grundlegende Entscheidung im Entwurf betrifft die Art der Anwendung. Für dieses Projekt kommen eine Webanwendung oder eine native Anwendung in Frage.

Auf Grund der Anforderung der D-School, dass die Anwendung auch außerhalb der Räumlichkeiten der D-School verfügbar sein muss (vgl. Anforderung F6), liegt eine Webanwendung nahe. Hinzu kommt, dass die Projektmitglieder bereits mit dem Umgang von Webseiten und deren Navigation vertraut sind. Für die Erweiterbarkeit stellt dies ebenfalls einen enormen Vorteil dar, da ein zentrales System gewartet werden kann und neue Funktionen einfach eingespielt werden können. Zudem dreht sich das Projekt um das Thema Dokumentation, bei welchem oft dazu geneigt wird, es aufzuschieben. Eine Webseite senkt hier die Hemmschwelle und umgeht die Notwendigkeit einer Verteilung und Installation eines Programms.

Eine Trennung der Applikation in Frontend und Backend erlaubt eine klare Funktionstrennung. Für Webapplikationen befindet sich das Backend auf Serverseite und das Frontend im Browser auf Clientseite. Die klassischen Aufgaben der beiden Teile sind:

Backend-Aufgaben

- Sammeln und zur Verfügung stellen von Daten
- Kommunikation mit anderen Servern
- Validierung eingehender Daten vom Client
- Speicherung von Daten
- Business-Logik

Frontend-Aufgaben

- Kommunikation mit dem Backend zur Daten-Synchronisation
- Visualisierung der Daten
- Interaktion mit dem Nutzer
- Feedback an den Nutzer

2.1.2 Generelle Architektur von Webapplikation-Backends

Die grundlegende Architektur beim Bau einer Webanwendung hängt bei der Wahl eines Webframeworks vom Stil dieses Frameworks ab. Die Verwendung eines Webframeworks erlaubt die Konzentration auf die Umsetzung der Webschnittstelle und beschleunigt die Programmierung von Webanwendungen durch das Forcieren von Prinzipien wie "Don't repeat yourself" (vgl. [HT99, p. 23]) oder "Convention over Configuration" (vgl. [Zyl08, p. 3]). Der Architekturstil, den die meisten Frameworks verwenden, ist das *Model-View-Controller* (MVC)-Pattern¹ (vgl. [GHJV95, p. 14]).

Das MVC-Pattern dient der Abstraktion der Datenhaltung und Business-Logik von der Präsentation. Im Kontext einer Webanwendung ist die Rollenverteilung wie folgt:

Model: Die Interaktion mit der Datenbank und Ausführung der Anwendungs-Logik.

View: Generierung von verschiedenen Darstellungen für die Daten, in der Regel im HTML, JSON² oder XML³ Format.

Controller: Annahme von HTTP-Anfragen und Verarbeitung dieser zu Anweisungen an das Model. Erzeugen einer Antwort mit Hilfe eines Views auf Basis des veränderten Models.

2.1.3 REST und CRUD

Bei der Umsetzung von HTTP-Schnittstellen hat sich der *Representational State Transfer* (REST)-Stil als gängige Alternative zu anderen Konzepten wie SOAP⁴ etabliert. Dieser legt ein URL-Schema für die Standardinteraktionen auf Ressourcen fest und verwendet dazu die verschiedenen HTTP-Operationen. Diese Aktionen umfassen *Create*, *Read*, *Update* und *Delete*, die als CRUD bezeichnet werden.

¹Aufgrund der allgemein üblichen Bezeichnung Model, View und Controller für die Teile einer MVC-Architektur werden diese anstatt der entsprechenden Begriffe Modell, Präsentation und Steuerung in dieser Arbeit verwendet.

²JavaScript Object Notation, <http://tools.ietf.org/html/rfc4627> (Zugriff 05.06.13)

³Extensible Markup Language, <http://www.w3.org/TR/REC-xml/> (Zugriff 18.06.13)

⁴Simple Object Access Protocol, <http://www.w3.org/TR/soap/> (Zugriff 28.06.13)

2.2 Datenmodellierung

In diesem Abschnitt sollen die Besonderheiten der in Project-Zoom verwendeten Datenmodelle erläutert werden. An ausgewählten Modellen werden die Besonderheiten der Umsetzung erklärt.

2.2.1 Überblick

In der Abbildung 2.1 ist ein Auszug aus dem Datenmodell für Project-Zoom gegeben. Ein komplettes Datenmodell ist im Anhang, in der Abbildung C.1, zu finden. Gezeigt sind die verschiedenen Datenklassen und ihre Assoziationen.

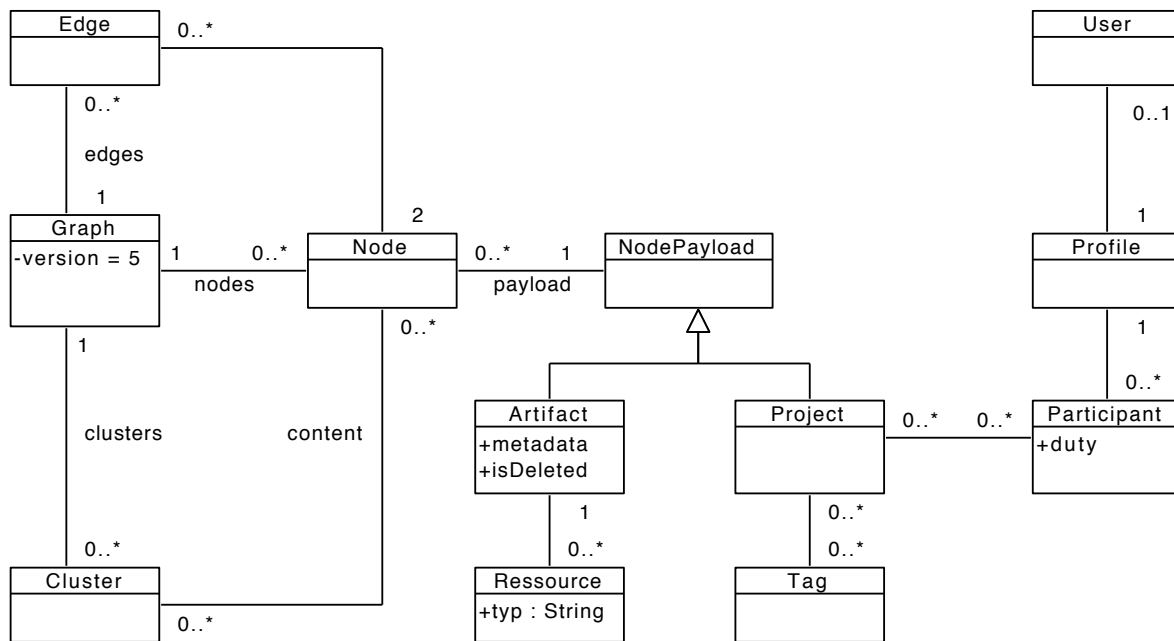


Abbildung 2.1 – Ausgewählte Datenmodelle und ihre Assoziationen

2.2.2 Datenmodell für Graphen

Ein Graph wurde als gerichteter Graph mit jeweils einer Liste für Knoten, Kanten und Cluster modelliert. Knoten und Kanten entsprechen denen aus der Graphentheorie bekannten Strukturen. Als konzeptionelle Referenz für den Aufbau eines Graphen dient [CLRS01, p. 531]. Cluster wurden eingeführt, um Nutzern die Möglichkeit zu geben, Knoten zu gruppieren und Phasen-Annotationen hinzuzufügen. Ein Objektdiagramm eines Graphen zeigt die Abbildung C.2 im Anhang.

In der Datenbank sind für den *Payload* eines Knotens nur Referenzen abgelegt. Der Payload beschreibt dabei den eigentlichen Inhalt eines Knotens, z.B. ein Projekt oder ein Artefakt. Wenn

der Graph über die REST-Schnittstelle ausgeliefert wird, dann wird diese Referenz auf den *NodePayload* mit dem eigentlichen Inhalt des Knotens ersetzt. Dies verringert den Speicherbedarf des Graphs in der Datenbank und verhindert das mehrfache Speichern des Payloads.

Versionierung

Um die Veränderungen an einem Graph nachverfolgen zu können, wurde eine Versionierung für Graphen eingeführt. Dadurch, dass nur Referenzen auf den Inhalt eines Knotens gespeichert werden, ist die Größe eines Graphens vergleichsweise gering. Aus diesem Grund ist es möglich, alle Versionen eines Graphens in der Datenbank abzulegen. Zur Identifizierung bekommt ein Graph eine Gruppe (*group*) zugewiesen. In einer Gruppe liegen die verschiedenen Versionen eines Graphen. Zusammen mit der Versionsnummer *version*, welche bei jedem Update inkrementiert wird, bildet die Gruppe einen Schlüssel.

Updaten eines Graphen

Um ein Graphen-Update ausführen zu können, wurde die Spezifikation von JSON-Patch⁵ zur Veränderung von JSON-Objekten gewählt. Diese definiert eine JSON-Struktur, um ein gegebenes Objekt durch die Anwendung einer geordneten Folge von Änderungen in einen neuen Zustand zu überführen. Das Patch-Objekt ist ein JSON-Array, dessen Elemente als Operationen auf das zu verändernde JSON-Objekt angewendet werden sollen. Die definierten Operationen sind das Hinzufügen, Testen, Entfernen, Ersetzen, Verschieben und Kopieren von Attributen.

Alternativ bestünde die Möglichkeit, bei jeder Änderung den kompletten, geänderten Graphen zu senden. Da es sich hier um kleine Objekte handelt, wäre der Mehrbedarf an Bandbreite vertretbar. Für den Mehrnutzerbetrieb ist allerdings die Verwendung von JSON-Patch vorteilhaft, denn durch das explizite Senden der Änderungen ist ein Zusammenführen gleichzeitiger Änderungen von unterschiedlichen Quellen einfacher. Im Hinblick auf ebendiesen Vorteil wurde in Project-Zoom JSON-Patch verwendet. Die Implementierung im Frontend ist in [Rze13] näher beschrieben.

2.2.3 Datenmodell für Artefakte

Artefakte werden durch externe Konnektoren gefunden und anschließend durch den Backend-Core in der Datenbank abgelegt, wo alle Informationen enthalten sind, die charakteristisch für das jeweilige Artefakt sind. Eine Besonderheit zeigt sich im Feld *metadata*. Hier können alle Informationen eines Konnektors gespeichert werden, die keinem allgemeinen Feld zugeordnet

⁵JavaScript Object Notation Patch, <http://tools.ietf.org/html/rfc6902> (Zugriff 15.06.13)

werden können. Durch diese Speicherung von Informationen, die zwar im Moment für die Anwendung irrelevant sind und nicht von allen Konnektoren geliefert werden können, vermeidet man einen Informationsverlust. Bei der Weiterentwicklung der Anwendung können dann Informationen aus diesen Metadaten extrahiert und als eigenständige Felder verwendet werden.

Zu einem Artefakt sind verschiedene *Ressourcen* abgelegt. Eine Ressource bezieht sich hierbei auf die digitale Repräsentation eines Artefakts. Das heißt, dass es zu jedem Ressourcenobjekt in der Datenbank eine zugehörige Datei im Dateisystem gibt. Beispielsweise ist die originale Datei eine Ressource, generierte Thumbnails sind weitere. Das Feld *typ* beschreibt die Art der Ressource und erlaubt einem Client die differenzierte Behandlung unterschiedlicher Typen.

2.2.4 Datenmodell für Nutzer

Neben Artefakten können durch einen externen Konnektor auch Profile gesammelt werden. Diese Profile bestimmen die Zugriffsrechte eines Nutzers. Der mit einem Profil verknüpfte User enthält die Anmeldeinformationen, die für eine Authentifizierung notwendig sind. Ein User wird erst dann zu einem Profil assoziiert, wenn sich ein Nutzer mit der im Profil angegebenen E-Mail-Adresse für Project-Zoom registriert. Aus diesem Grund ist es auch nur Nutzern erlaubt sich zu registrieren, für die bereits ein Profil existiert.

Die Abspaltung des Users als Container für die Authentifizierungsinformationen vom Profil hat weiterhin den Vorteil, dass diese Informationen nicht über einen Konnektor geändert werden können. Wäre das der Fall, so könnte ein Angreifer die Log-In-Informationen für Project-Zoom durch das Verändern von Daten in einem Konnektor so anpassen, dass er Zugriff auf den Nutzer-Account bekommt.

2.3 Data-centric Design

In diesem Kapitel erfolgt eine Erläuterung der grundlegenden Architektur der Anbindung eines persistenten Speichers im Core von Project-Zoom. Das gewählte Design hat vorrangig Auswirkungen auf die Art und Weise, wie der Code für Datenmodelle geschrieben wird, erstreckt sich aber als Betrachtungsweise über die gesamte Architektur. Zuerst wird das verwendete data-centric Design näher beschrieben und mit alternativen etablierten Formen der Datenbankanbindung verglichen. Im Kapitel 3.2 wird näher auf die Umsetzung des Designs in Project-Zoom eingegangen.

2.3.1 Motivation

Das JSON Coast-To-Coast Design als Umsetzung des data-centric Designs stellte erstmals zusammenhängend und mit Beispielen unterlegt Pascal Voitet in [Voi13] dar. Voitet ist Mitwirkender am Open-Source-Projekt *Play*⁶ und Autor von Scala-Bibliotheken (*play-reactivemongo*⁷, *play-autosource*⁸).

Ein Backend für eine Webapplikation ist zunehmend eine Verbindung zwischen verschiedenen anderen Backends und Frontends. Diese Entwicklung geht auf die Bereitstellung von APIs für viele große und kleine im Web verfügbare Dienste zurück. Anwendungen, welche Daten vorrangig aus anderen Quellen aggregieren, sind zum großen Teil mit der Datenmanipulation beschäftigt. Hier bietet sich ein data-centric Design an. Grundlegend für das Design sind dabei folgende Entwicklungen: Zum einen das Aufstreben von NoSQL-Datenbanken und zum anderen asynchrone Datenbanktreiber mit einfachen Methoden zur JSON-Manipulation.

NoSQL

Die Entwicklung der Datenbankmanagementsysteme beschränkte sich viele Jahre auf die Optimierung und Verbesserung bestehender relationaler Datenbankmodelle. Im Jahr 1998 kam der Term *Not-only SQL* (NoSQL) auf (vgl. [LM10]). Heute gibt es verschiedene etablierte Datenbanken, die keine reinen SQL-Datenbanken mehr sind, wie beispielsweise MongoDB⁹, Apache CouchDB¹⁰ und Apache Cassandra¹¹.

NoSQL-Datenbanken sind vorteilhaft für ein data-centric Design, da zur Umsetzung eines Datenflusses (vgl. Abbildung 2.2) eine Ähnlichkeit des Datenbank- und Clientformats vorliegen muss. Diese Ähnlichkeit liegt zum Beispiel bei BSON und JSON vor und verhindert somit die Notwendigkeit einer Konvertierung.

Datenbanktreiber

Für die Anbindung von relationalen Datenbanken gibt es in Java die Java Database Connectivity (JDBC). Diese Schnittstelle abstrahiert über Datenbanken und deren Treiber, indem eine einheitliche API angeboten wird. Ausgerichtet ist JDBC auf relationale Datenbanken (vgl. [Ree00]).

⁶Play Framework, <http://playframework.com> (Zugriff 29.06.13)

⁷ReactiveMongo, <https://github.com/zenexity/Play-ReactiveMongo> (Zugriff 15.06.13)

⁸Play Autosource, <https://github.com/mandubian/play-autosource> (Zugriff 15.06.13)

⁹MongoDB, <http://mongodb.org> (Zugriff 17.06.13)

¹⁰Apache CouchDB, <http://couchdb.apache.org> (Zugriff 17.06.13)

¹¹Apache Cassandra, <http://cassandra.apache.org> (Zugriff 17.06.13)

Um NoSQL-Datenbanken anzubinden, benötigt man, wie bei JDBC, einen eigenen Treiber. Der Unterschied ist, dass es hier keine Abstraktionsebene über verschiedene NoSQL-Datenbanken gibt. Dies liegt vorrangig an der sich stark unterscheidenden Struktur der einzelnen Speichersysteme¹². Die unterschiedlichen NoSQL-Datenbanken sind jeweils speziell auf eine bestimmte Aufgabe ausgerichtet, wie z.B. auf Durchsatz, verteilte Umgebungen oder flexible Datenschemas.

ReactiveMongo ist ein Beispiel für einen asynchronen Datenbanktreiber für MongoDB. Die Vorteile eines asynchronen Treibers liegen auf der Hand: Für jede synchrone Datenbankabfrage wird normalerweise ein Thread verwendet, der bis zur Antwort blockiert ist. Bei mehreren Datenbankabfragen pro Request werden bei Last viele Threads benötigt, um Datenbankabfragen auszuführen. Asynchrone Treiber umgehen dieses Problem, indem sie Threads, welche Datenbankabfragen ausführen, nicht blockieren. Für dieses Konzept ist es notwendig Platzhalter einzuführen. Diese ersetzen das Ergebnis, solange es noch nicht vorhanden ist.

Quelltext 2.1 – Funktionssignatur für asynchronen Datenbankzugriff

```
def findById(bid: BSONObjectID): Future[Graph] = {  
    collection.find("_id" -> bid)  
}
```

Im Quelltext 2.1 zeigt die Funktionssignatur den Rückgabebetyp *Future[Graph]*. *Future* ist hierbei ebendieser Platzhalter für ein Ergebnis, welches noch nicht existiert. Futures basieren dabei auf einem Konzept, welches in [BH77] beschrieben ist. Um auf das Ergebnis eines Futures zu reagieren, können verschiedene *Callbacks* festgelegt werden.

Die Verwendung eines asynchronen Treibers ist vor allem im Hinblick auf das zu implementierende, asynchrone Eventsystem sinnvoll und von Vorteil.

Manipulation von JSON

Das data-centric Design basiert auf der Idee der direkten Manipulation von Daten. Für eine Umsetzung des data-centric Designs mit Hilfe von JSON benötigt man somit eine Bibliothek, welche die Möglichkeit bietet, JSON zu Transformieren und zu Validieren.

Eine solche Bibliothek, welche das Validieren, Transformieren und Serialisieren erlaubt, ist Play-JSON. Ein Beispiel, wie solch eine Transformation und Validierung aussieht, ist im Anhang im Quelltextauszug A.1 zu sehen. Dort wird ein JSON-Objekt erzeugt, welches anschließend zuerst transformiert wird, indem das Attribut *role* angehängen wird, und danach validiert wird. Bei der Validierung wird überprüft, ob der Nutzer über 18 Jahre alt ist.

¹²Die gängigsten Formate sind dokumentenbasierte Speicher oder einfache Schlüssel-Wert-Speicher (vgl. [Pok11]), es gibt jedoch noch eine Reihe weiterer Speichersysteme.

2.3.2 Idee hinter data-centric Design

Das Ziel im data-centric Design ist die direkte Manipulation von Daten. Der Fokus des Designs liegt dabei sowohl auf der Datenmanipulation als auch auf dem Datenfluss. Wie bereits im Abschnitt 2.3.1 beschrieben, sind Webapplikation-Backends immer häufiger Knotenpunkte in einem mehrere Server durchlaufenden Datenfluss. Ein Teil dieses Datenflusses, die Kommunikation des Backends mit der Datenbank und dem Client, ist in Abbildung 2.2 dargestellt.

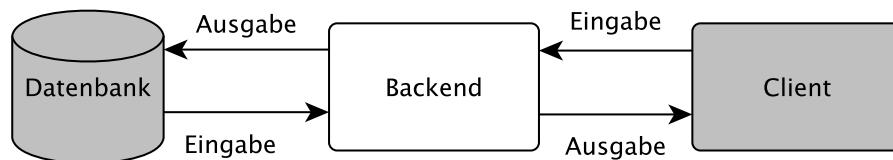


Abbildung 2.2 – Datenfluss zwischen Datenbank, Backend und Client ¹³

In diesem Datenfluss soll es keine implizite Umwandlung der Daten in Objekte der Programmiersprache¹⁴ geben. Für ein einfaches Durchleiten der Daten aus der Datenbank zum Client oder von einer anderen WebRessource zum Client ist eine Umwandlung nicht notwendig oder sinnvoll.

Im Gegensatz zu anderen Modellen ist die Abstraktion von der Datenquelle kein Ziel des data-centric Designs. Das Festlegen auf eine Datenquelle bei der Entwicklung eines Systems erlaubt die Verwendung aller Funktionen, die diese Datenquelle zur Verfügung stellt. Beispielsweise erlaubt MongoDB die Verwendung sogenannter *Capped Collections*¹⁵, welche auf hohen Durchsatz optimiert sind. Diese verhalten sich wie eine zirkuläre Queue mit begrenzter Größe. Viele Datenbanken besitzen spezielle, stark angepasste Funktionen, auf die jedoch bei der Verwendung von allgemeinen Datenbank-APIs (z.B. JDBC) meist nicht zugegriffen werden kann. Ein Nachteil dieser engen Kopplung ist der schwierigere Austausch der Datenquelle.

Eine Deserialisierung der Daten in Objekte ist für komplexe Business-Logik von Vorteil. Statische Datenmodelle sollten immer dann verwendet werden, wenn die Manipulation der Daten kompliziert wird. Hier kommt der Nachteil von dynamischen Speicherstrukturen zum Vorschein, denn sobald es nicht mehr darum geht ein JSON-Objekt zu transformieren oder zu validieren, sondern auf einzelne Attribute zuzugreifen, muss dieses Attribut immer einzeln deserialisiert werden. Das schließt die Berücksichtigung der Fehlerfälle, wie dem Fehlen des Attributes oder

¹³vgl. Pascal Voitet in [Voi13]

¹⁴Es ist bekannt, dass die Repräsentation der Daten des Datenflusses ebenfalls in Objekten der Programmiersprache erfolgt (beispielsweise JSON-Objekte). Mit dem Ausdruck ist die Überführung der Daten aus einem allgemeinen Modell (z.B. JSON) in eine spezifische Objektdarstellung (z.B. ein Objekt einer User-Klasse) gemeint.

¹⁵Weitere Informationen zu Capped Collections z.B. auf <http://docs.mongodb.org/manual/core/capped-collections/> (Zugriff 20.06.13)

dem Vorhandensein des falschen Typs, mit ein. Aus diesem Grund ist an dieser Stelle eine Umwandlung in ein Objekt der Programmiersprache vor der Verwendung sinnvoll. Ebenfalls notwendig ist diese Konvertierung bei der Nutzung der meisten externen Bibliotheken.

Nach Voitet sollten statische Datenmodelle also nicht vergessen, sondern nur dann genutzt werden, wenn sie notwendig sind. Einfache und dynamische Strukturen sollten so oft es geht erhalten bleiben (vgl. [Voi13]).

2.3.3 Abgrenzung zur objektrelationalen Abbildung

Das *Object-relational Mapping* (ORM) wird auch als *All-Model-Approach* bezeichnet. Hierbei erfolgt die Kommunikation mit der Datenbankschnittstelle über Objekte der Programmiersprache. Abfragen an die Datenbank resultieren in Objekten der Programmiersprache. Der Fokus der objektrelationalen Abbildung liegt in der Überführung von Objekten der objektorientierten Programmierung in ein relationales, tabellenorientiertes Schema (vgl. [Amb03]).

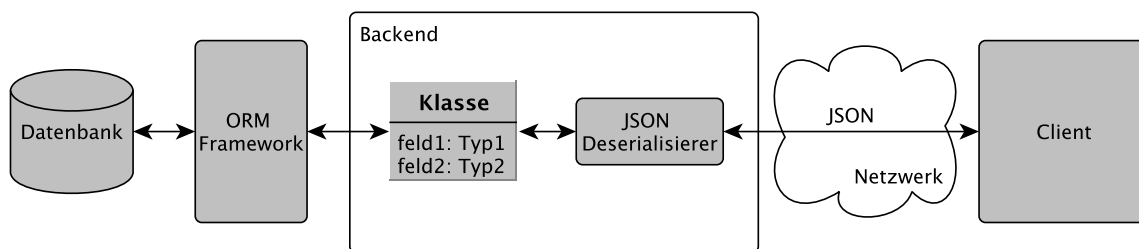


Abbildung 2.3 – Datenfluss einer Applikation bei Verwendung eines ORM-Frameworks ¹⁶

Der wichtigste Vorteil eines Frameworks für die objektrelationale Abbildung liegt in der Abstrahierung von der Datenbank. Für die Business-Logik gibt es keinen Unterschied zwischen Objekten aus der Datenbank und Objekten der Programmiersprache. Die Datenbank kann meist ohne Änderungen an der Applikation ausgetauscht werden.

Objektrelationale Unverträglichkeit

Ein ORM bringt konzeptionelle Probleme mit sich. Diese werden in der Literatur (vgl. [IBNW09]) in der Regel als *Object-relational Impedance Mismatch* (ORIM) bezeichnet. Die objektrelationale Unverträglichkeit resultiert aus den Unterschieden im zugrundeliegenden Konzept, sowie aus Differenzen in der Darstellung und den Erwartungen des Entwicklers (vgl. [Bow10]). Folgende

¹⁶vgl. Pascal Voitet in [Voi13]

Herausforderungen ergeben sich für die praktische Arbeit mit einem ORM:

- Die Abbildung der Assoziationen zwischen den Objekten ist komplex. Es müssen One-to-One, One-to-Many und Many-to-Many Abhängigkeiten ins relationale Schema übertragen werden.
- Ein weiterer zentraler Punkt ist die Abbildung von Hierarchien und Vererbungen aus der objektorientierten Programmierung in das relationale Datenmodell.
- Innerhalb des ORM muss es ein Objekt-Caching geben. Dieses ist notwendig um die Illusion zu erzeugen, dass es sich um ein Objekt der Programmiersprache handelt. Denn wird ein Objekt mehrmals aus der Datenbank abgefragt, so muss ein referenzgleiches Objekt zurückgegeben werden (vgl. [HVE03]).
- Der Entwickler muss die ORM-Limitierungen, die sich aus dem ORIM ergeben, akzeptieren. Die Probleme müssen somit bekannt sein, um sie bereits bei der Datenmodellierung zu umgehen (vgl. [Atw06]).

Ted Neward, Autor von „The Busy Java Developers Guide to Scala“ [New08], macht den frühen Erfolg von objektrelationalen Abbildungen und die Erleichterungen, welche die ersten ORMs mit sich brachten, dafür verantwortlich, dass zum Teil ORMs verwendet werden, ohne deren Limitierungen zu beachten. Das wiederum führt zu einem erheblichen Mehraufwand an Zeit und Energie für die Erhaltung und Anpassung an alle Nutzungsfälle (vgl. [Atw06]).

Unterschiede zum data-centric Design

Das data-centric Design ist viel stärker an die Datenbank gebunden und deren Austausch ist deutlich schwieriger. Dies sorgt dafür, dass keine Illusion einer abstrahierten Datenbank entsteht. Da das data-centric Design ein Ansatz mit funktionalem Hintergrund ist, sind die Datenstrukturen zur Laufzeit in der Regel unveränderbar¹⁷. Ein Caching zur Sicherung der Objektidentität ist daher nicht notwendig. Die Objektgleichheit ist in diesem Fall über Wertgleichheit definiert (vgl. Scala Case-Klassen in [Sca08]). Im Gegensatz zum ORM muss sich der Entwickler beim data-centric Design selbst mit der Darstellung und Umwandlung seiner Daten in das Datenbankformat beschäftigen. Durch die Verwendung einer objekt- oder dokumentenbasierten Datenbank ist diese Umwandlung einfacher und mit weniger Problemen behaftet, als eine Konvertierung in ein relationales Schema.

¹⁷Die hier beschriebene Unveränderbarkeit bezieht sich auf die in der funktionalen Programmierung übliche Bedeutung. Die Struktur eines Objektes ist zur Entwurfszeit beliebig veränderbar. Zur Laufzeit können den Attributen eines Objektes nach Erstellung dessen keine neuen Werte mehr zugewiesen werden. Der Zustand des Objektes ist somit unveränderbar (vgl. [WP10, p. 181]).

2.3.4 Abgrenzung zu Datenzugriffsobjekten

Implementierungen von Datenzugriffsobjekten, in der englischsprachigen Literatur *Data Access Object* (DAO) oder *Data Access Procedure*, basieren auf dem *Data Access Object*-Pattern (vgl. [ACM03]). Das Pattern hat das Ziel, den Datenzugriff und die Datenmanipulation in einen separaten Layer auszulagern. Daraus ergibt sich eine Kapselung der Datenbankzugriffe an einem Ort der Applikation. Bei der Umsetzung gibt es die Möglichkeit, ein DAO zur Abstraktion der Datenbank als Ganzes zu verwenden oder für jede einzelne Datenbankstruktur ein DAO anzulegen.

Der Hauptfokus liegt in der Abstrahierung der Datenbank. Bei Änderungen an der Datenquelle müssen in der Regel nur die Datenzugriffsobjekte angepasst werden, denn der Anwendungscode bleibt im Allgemeinen unberührt. Die Serialisierung der Daten von den Strukturen der Anwendung in Strukturen der Datenbank ist nicht Aufgabe des DAO, sondern des *Data Transfer Object* (DTO)¹⁸.

Im DAO findet über einen Datenbanktreiber direkter Zugriff auf die Datenbank statt. Man kann das Datenzugriffsobjekt somit als Bindeglied zwischen Datenbank und Applikation sehen. Durch die direkte Kommunikation mit der Datenquelle können alle unterstützten Funktionen ausgenutzt werden.

Sowohl DAO als auch ORM haben die Abstrahierung der Datenbank als Ziel. Im Gegensatz zum DAO geht das ORM jedoch noch einen Schritt weiter und abstrahiert nicht nur über die Funktionen, welche durch die Datenbank zur Verfügung gestellt werden, sondern auch über die innere Struktur der Datenquelle. In der Regel bauen ORM-Implementierungen auf dem Data Access Object-Pattern und dem Data Transfer Object-Pattern auf.

Unterschiede zum data-centric Design

Das DAO-Pattern enthält, ähnlich dem data-centric Design, keine Serialisierung oder Deserialisierung von Objekten. Beide Designs können sehr gut zusammen verwendet werden. Der Datenfluss kann somit durch ein DAO realisiert werden, wodurch die Vorteile des Datenzugriffsobjektes, wie zum Beispiel eine stärkere Kapselung, genutzt werden. Voitet benutzt in seinen Beispielen kein DAO, sodass sich die Datenbankabfragen über den kompletten Quellcode verteilen. Während das bei Beispielapplikationen oder kleineren Anwendungen kein Problem darstellt, geht bei größeren Anwendungen mit vielen verschiedenen Datenmodellen schnell die Übersichtlichkeit verloren. Die Verwendung eines Datenzugriffsobjektes führt zu einem *Single Point of Responsibility*¹⁹ und stärkt somit die Struktur des gesamten Projektes.

¹⁸vgl. J2EE Patterns – Data Transfer Object <http://www.oracle.com/technetwork/java/transferobject-139757.html> (Zugriff 20.06.13)

¹⁹Single Responsibility Principle (SRP), vgl. [FFBS04, p. 339]

2.3.5 Grenzen des data-centric Designs

Das data-centric Design eignet sich sehr gut für *Create-Read-Update-Delete* (CRUD)-Anwendungen. Dort liegt der Fokus auf der Datenmanipulation, was zum Grundsatz des Designs von Project-Zoom passt.

Es erfolgt eine explizite Bindung an das Datenmodell. Durch die Nähe zur Datenquelle können spezifische Fähigkeiten einer Datenquelle genutzt und Anfragen effektiver gestaltet werden. Die dynamischen Strukturen des Datenmodells ermöglichen eine leichte Veränderbarkeit des Modells sowie einfache Verknüpfungen zwischen Daten.

Grenzen sind dem data-centric Design vor allem durch die Business-Logik gesetzt. Durch die dynamischen Strukturen im Design ist der Code meist komplizierter und länger.

Ohne die Verwendung eines DAO gibt es in diesem Design keinen Single point of responsibility für Datenbankzugriffe. Das erschwert wiederum Änderungen am Schema, da diese zu unvorhergesehenen Problemen an anderen Stellen im Anwendungs Quelltext führen können. Aus diesem Grund sollten bei diesem Design ein oder mehrere Datenzugriffsobjekte verwendet werden. Ein weiterer Vorteil bei der Verwendung eines DAO ist die deutlich leichtere Austauschbarkeit der Datenquelle durch die Zentralisierung der Datenbankabfragen im Datenzugriffsobjekt.

2.4 Eventsystem

In der Anforderung F2 wurde festgelegt, dass die von den Teams erstellten Dokumente aus der jeweiligen Box für die Studierenden in Project-Zoom zur Verfügung stehen sollen. Diese Verfügbarkeit soll schnellstmöglich nach Hinzufügen einer neuen Datei zur Box gegeben sein (siehe Anforderung NF2). Hierfür ist es sinnvoll ein asynchrones System zu bauen. Neben der Anbindung von Box ist für die Zukunft auch der Zugriff auf Facebook, Dropbox und Netzwerkdateisysteme vorgesehen. Es zeigt sich, dass ein paralleles, unabhängiges Abfragen der einzelnen Dienste notwendig ist.

Ein Eventsystem basiert auf dem Prinzip des Abonnierens und Veröffentlichens von Nachrichten und modelliert eine Eine-zu-Viele-Relation. Ein sogenannter *Publisher* sendet seine Events, welche anschließend vom *Dispatcher* an die *Subscriber* verteilt werden. Die Subscriber können dieses Event schließlich abarbeiten (vgl. [GHJV95, p. 327]). Das Publisher-Subscriber Modell kann sowohl synchron als auch asynchron implementiert werden. Bei einer asynchronen Umsetzung wird ausgenutzt, dass der Sender beim Emittieren einer Nachricht sofort sequentiell weiterarbeiten kann und die Bearbeitung des Events nicht abwarten muss.

Die parallele Abarbeitung von verschiedenen Aufgaben ist ein wichtiger Bestandteil von Project-Zoom. Ein Beispiel hierfür ist das Erzeugen von Thumbnails, was bei größeren Dateien eine

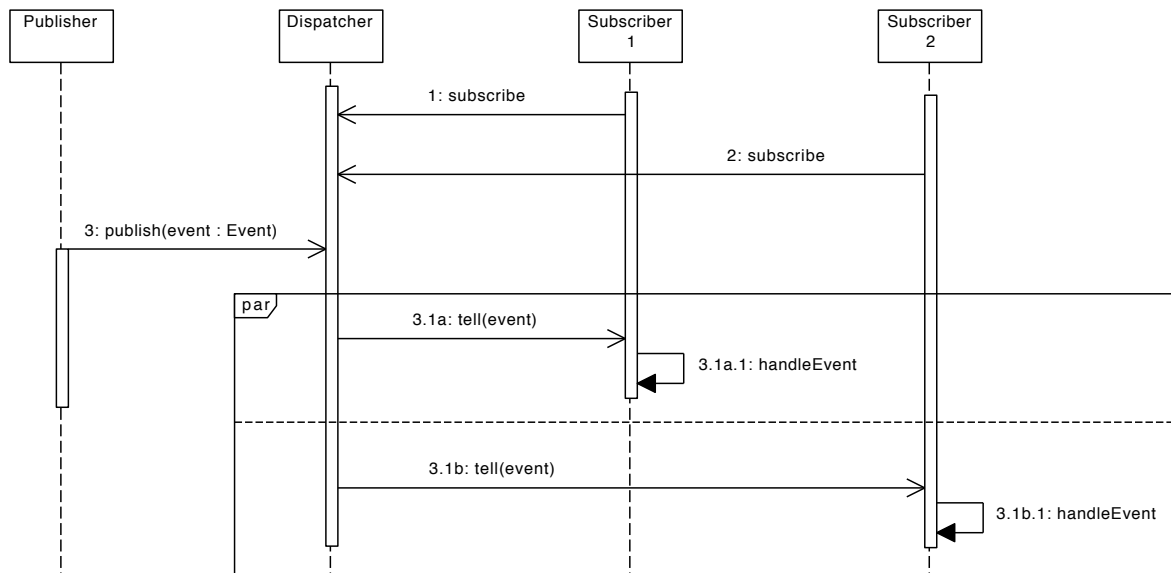


Abbildung 2.4 – Funktionsweise des Eventsystems an einem Beispiel

gewisse Zeit in Anspruch nehmen kann. Diese Thumbnails werden für jedes durch einen Konnektor gefundene Artefakt erzeugt. Während der Thumbnailerzeugung muss die Webapplikation weiterhin Client-Anfragen beantworten können, weswegen die Thumbnailgenerierung ausgelagert wurde.

Ebenfalls von Bedeutung ist die Ausfallsicherheit. Der Ausfall einer einzelnen Komponente darf die Stabilität des Gesamtsystems nicht beeinträchtigen. Eine falsch konfigurierte OpenOffice-Anbindung, welche der Thumbnailgenerator benötigt (vgl. [Brä13]), darf, selbst wenn der Thumbnailerzeuger nicht mehr in der Lage ist zu arbeiten, das System nicht zum Stillstand bringen.

Ein weiterer Vorteil ist die Flexibilität, die durch ein Publisher-Subscriber-Modell erreicht wird. Wenn eine neue Komponente eingebunden werden soll, so kann sich diese auf bereits emittierte Nachrichten abonnieren und somit auf Ereignisse im Gesamtsystem reagieren.

In Abbildung 2.4 ist der Aufbau eines Eventsystems zu sehen. Der Dispatcher ist die zentrale Komponente im Eventsystem, denn er verwaltet die Abonnements und verteilt die Events an die Subscriber. Diese Verantwortung stellt eine Schwachstelle im Eventsystem dar, es existiert ein *Single Point of Failure*. Ein Ausfall kann durch die Vervielfältigung des Dispatchers vermieden werden, wobei eine gemeinsame Datenbasis für alle Dispatcher notwendig ist.

3 Umsetzung des Backends

Die Umsetzung der gewählten Architektur und Konzepte in Project-Zoom ist Thema dieses Kapitels. Zuerst wird auf die Technologieauswahl eingegangen und die Implementierung des data-centric Designs kurz erläutert. Anschließend wird die Realisierung der Anbindung der externen Komponenten an das Eventsystem am Beispiel verdeutlicht. Den Abschluss bilden Auswertung sowie Betrachtung der Umsetzung der Anforderungen.

3.1 Technologieauswahl

3.1.1 Scala im Web: Play Framework

Die Auswahl der Programmiersprache wurde vor allem von dem Gedanken geprägt, eine Sprache zu finden, die einerseits auf benötigte Bibliotheken zugreifen kann und andererseits eine kurze Einarbeitungszeit benötigt. Für die Arbeit mit Thumbnails hat sich die Bibliothek *Apache Tika*¹ als besonders wertvoll erwiesen. Nähere Informationen zu dieser Bibliothek finden sich in [Brä13]. Da Apache Tika in Java geschrieben wurde, musste die Wahl auf eine Sprache fallen, welche in der Lage ist, Java-Bibliotheken anzusprechen.

Um die verschiedenen Sprachen und ihr jeweiliges Webframework zu evaluieren, wurde sich mit den in Tabelle 3.1 aufgelisteten Kombinationen näher beschäftigt. Dazu wurde von den Backend-Entwicklern jeweils eine kurze Hello-World-Anwendung aufgesetzt. Dadurch konnte festgestellt werden, ob der Arbeitsablauf, der Aufwand und der zu erzeugende Quellcode angemessen sind. Neben dieser praktischen Kurzevaluierung wurden einige Fakten zusammengetragen, um die Entscheidung zu erleichtern. Die Kriterien Asynchronität und Zustandslosigkeit sind dabei vor allem im Kontext eines asynchronen Eventsystems von Interesse. Bezüglich der Indikatoren *Adoption Ready*² und *Importance*³ befinden sich alle Frameworks auf einem ähnlichen Niveau.

¹Apache Tika, <http://tika.apache.org> (Zugriff 29.06.13)

²Umfrage von InfoQ mit 1894 Teilnehmern auf <http://www.infoq.com/research/jvm-web-frameworks> (Zugriff 26.06.13). Erfragt wurde eine Einschätzung der Reife des Frameworks. Der Stand bei Abruf findet sich in der Tabelle C.3 im Anhang.

³Entnommen aus der bereits erwähnten InfoQ-Umfrage. Erfragt wurde eine Einschätzung der Relevanz des Frameworks.

In die nähere Auswahl kamen Java, Groovy und Scala⁴ mit den jeweiligen MVC-Webframeworks Spring MVC⁵, Grails⁶ und Play. Im Gegensatz zu Spring und Grails ist Play ein zustandsloses, schlankes Framework, was eine ideale Voraussetzung für ein REST-Backend darstellt. Zusätzlich spielten die persönlichen Präferenzen der Entwickler eine Rolle, die bereits Erfahrung in der Kombination Scala und Play hatten.

Scala ist eine Programmiersprache, deren Syntax an Java orientiert ist. Programme, welche in Scala geschrieben sind, können bidirektional mit Java-Code interagieren und so auf den vollen Funktionsumfang der Java-Bibliotheken zurückgreifen. Dabei werden in der Sprache Konzepte der funktionalen mit gewohnten Elementen der objektorientierten Programmierung verknüpft.

	Java + Spring	Groovy + Grails	Scala + Play
Authentifizierung	Ja mit Spring-WS	Ja mit Authentication Plugin	Ja mit SecureSocial
Json-Unterstützung	Ja mit Jersey	Ja	Ja
Vorhandene Erfahrung im Entwicklerteam	Gering	Gut	Sehr Gut
Dokumentation	Gut	Gut	Gut
Asynchron	Nein	Nein	Ja
Zustandslos	Nein	Nein	Ja
<i>Adoption Ready</i>	85%	77%	71%
<i>Importance</i>	82%	76%	78%

Tabelle 3.1 – Vergleich von verschiedenen Webframeworks miteinander

3.1.2 MongoDB als Datenbanksystem

Während der Prototypen-Phase, in der sich die Idee der Umsetzung verfeinerte, wurde das bereits beschriebene Datenmodell aufgestellt. Ein wichtiger Aspekt, der für die Wahl der Datenbank entscheidend ist, sind die verschiedenen, anbindbaren, externen Systeme. Das beschriebene Datenmodell verwendet flexible Datenstrukturen zur Speicherung von zusätzlichen Informationen zu Artefakten. Dies muss also auch durch die Datenbank unterstützt werden. Weiterhin müssen die Graphstrukturen in der Datenbank abgelegt werden. Dafür eignen sich besonders Datenspeicher, welche strukturierte Daten zulassen. Aus diesen Gründen fiel die Entscheidung auf eine NoSQL-Datenbank, welche die benötigte Flexibilität ohne großen Aufwand ermöglicht (vgl. [SA12]).

⁴Nähere Informationen zu Scala und den Aspekten der Programmiersprache finden sich in [Ode11].

⁵Spring MVC, <http://www.springsource.org/features/modern-web> (Zugriff 26.06.13)

⁶Grails, <http://grails.org> (Zugriff 26.06.13)

Neben der Open-Source-Datenbank MongoDB stand Apache CouchDB als Alternative zur Verfügung. Die Wahl von MongoDB als persistenten Speicher für Project-Zoom fiel vorrangig aufgrund des sehr guten Datenbank-Treibers ReactiveMongo für Scala. Dieser erlaubt komplett asynchrone Datenbankzugriffe und gliedert sich deshalb perfekt in das asynchrone Play Framework ein.

Als dokumentorientierter Datenspeicher passt MongoDB sehr gut zu einer REST-Architektur und dem data-centric Design. Die Daten, abgelegt in der Datenbank, werden im Binary JSON (BSON)-Format gespeichert. Dieses BSON-Format ist eine binär encodierte Serialisierung von JSON-ähnlichen Dokumenten. BSON unterstützt die Repräsentation aller Datentypen, welche auch von JSON unterstützt werden, und erlaubt zusätzlich weitere Datentypen. So wurden die JSON-Datentypen unter anderem um die Unterstützung von Binärdaten und Datumsangaben erweitert (vgl. [Dir10]). Damit ähnelt die Datenform, die auf Clientseite verarbeitet wird, dem Format der Datenspeicherung in der Datenbank. Diese Konstellation erlaubt die Implementierung einer schlanken Schicht des Datenmodells, wie sie in dem Kapitel 2.3 erklärt wurde.

Zugriffsschutz für Datenbankobjekte

Eine Anforderung der D-School bestand in dem Schutz der Daten (vgl. Anforderung F1). Diesen Schutz gilt es für die Objekte in der MongoDB umzusetzen. Dabei soll verhindert werden, dass Informationen aus einem Projekt für Personen, die kein Geheimhaltungsvertrag für das Projekt unterschrieben haben, zugänglich sind. Studierende haben folglich nur Lesezugriff auf Projekte, an denen sie selbst teilgenommen haben oder die veröffentlicht⁷ wurden. Schreibzugriff erhält ein Student nur als in der Filemaker-Datenbank vermerkter Teilnehmer des Projektes.

Bei der Betrachtung der Struktur von Zugriffsbeschränkungen fällt auf, dass diese abhängig von der zugriffssuchenden Person, der Art des Zugriffes und der Ressource sind. Basierend auf dieser Erkenntnis wurde ein sogenannter *DBAccessContext* eingeführt. Für jeden Datenbankzugriff wird ein solcher Kontext benötigt. Mit Hilfe dessen können auf unterster Zugriffsebene, in den Funktionen *insert*, *update*, *remove* und *find*, die Berechtigungen überprüft werden.

Das Datenzugriffsobjekt für das jeweilige Datenmodell kann schließlich festlegen, wie mit Hilfe des *DBAccessContext* der Zugriff auf diese Methoden reguliert wird. Jedes dieser Datenzugriffsobjekte erbt von *SecuredDAO*. In dieser Klasse finden die Ausführung der Datenbankfunktionen und die Zugriffsbeschränkung statt. Dadurch, dass die Zugriffsbeschränkungen für die jeweiligen Datenbankaktionen einzeln festgelegt werden können, ist eine sehr feingranulare Differenzierung möglich.

⁷Aufgrund der fehlenden Einstellungsmöglichkeiten in der Filemaker-Datenbank gibt es in der aktuellen Implementierung nicht die Möglichkeit ein Projekt als öffentlich zu kennzeichnen.

Die Klasse *SecuredDAO* überprüft die Berechtigungen wie folgt:

insert: Beim Hinzufügen wird direkt kontrolliert, ob der User die Berechtigung hat, neue Elemente in das jeweilige Datenmodell einzufügen. Fehlt diese, wird die Aktion mit einem Fehler abgebrochen.

update: Eine Update-Aktion benötigt immer eine Suchanfrage und das eigentliche Update, welches auf die gefundenen Ergebnisse angewandt wird. Soll ein User nicht alle Objekte updaten können, so wird an die Suchanfrage die jeweilige Einschränkung angehängen. Damit wird verhindert, dass Objekte gefunden werden, auf die der User keinen Zugriff hat. Objekte, die nicht gefunden werden, können demnach auch nicht verändert werden.

find, remove: Für das Abfragen und Löschen ist eine Suchanfrage notwendig. Diese wird genauso verändert, wie dies für die *update*-Funktion beschrieben ist. Der Ergebnisraum wird somit schon während der Abfrage auf die dem User zugänglichen Objekte beschränkt.

3.1.3 Verwendete Bibliotheken

Die umfangreichsten Bibliotheken, welche im Project-Zoom Backend-Core verwendet werden, sind Akka⁸, Play und SecureSocial⁹. Nicht hier aufgeführt ist ReactiveMongo als Datenbanktreiber. Diese Bibliothek wurde bereits im Abschnitt 2.3.1 näher erläutert.

Akka stellt die Grundlage für das Play Framework dar. Die Bibliothek ermöglicht die Nutzung verschiedener Konzepte des asynchronen Programmierens. Das Ziel ist, das Schreiben von parallelen, fehlertoleranten und skalierbaren Anwendungen zu vereinfachen (vgl. [Akk12]).

Die sogenannten *Actors*¹⁰ der Bibliothek sind für dieses Projekt am relevantesten. Sie stellen eine Implementierung des Aktorenmodells dar. Aktoren sind abgeschlossene Einheiten, welche nur über Nachrichten kommunizieren. Dabei erfolgt die Abarbeitung der Nachrichten eines Aktors sequenziell, die Kommunikation mit anderen Aktoren hingegen asynchron. Dadurch können mehrere Nachrichten in unterschiedlichen Aktoren gleichzeitig abgearbeitet werden. Verschiedene Aktoren teilen sich nur über Nachrichten ausgetauschte Variablen. Damit das System also threadsicher arbeitet, müssen diese Nachrichten threadsicher sein. In Scala ist es deshalb üblich, für Nachrichten *Case Classes*¹¹ zu verwenden. Diese sind von vornherein unveränderbar und somit threadsicher.

Neben Actors finden auch *Agents* in Project-Zoom Verwendung. Ein Agent bildet eine Kapselung um einen Zustand. Dieser Zustand kann unverzüglich synchron gelesen und asynchron

⁸Akka Framework, <http://akka.io> (Zugriff 20.06.13)

⁹SecureSocial, <http://securesocial.ws> (Zugriff 20.06.13)

¹⁰Die Idee von Aktoren wurde erstmals in [HBS73] beschrieben.

¹¹Weiterführende Informationen zu Case Classes auf <http://www.scala-lang.org/node/107> (Zugriff 26.06.13)

überschrieben werden. Bei einem Update wird dem Agent eine Funktion übergeben, welche den neuen Zustand des Agents berechnet. Einen Agent kann man zum Beispiel verwenden, um den Zustand zwischen verschiedenen Actors zu teilen.

Play Framework 2.0 ist die Weiterentwicklung und Portierung eines früheren Java Webframeworks nach Scala. Es ist zwar in Scala programmiert, kann aber sowohl mit Java als auch Scala als Backend-Sprache verwendet werden. Play liegt eine MVC-Architektur zugrunde. Hierbei werden in jedem Controller sogenannte *Actions* definiert und im View verschiedene *Templates* angelegt. Der Ablauf einer Anfrage an den Webserver verläuft wie folgt:

1. Der standardmäßige HTTP-Router leitet die Anfrage an eine Action weiter. Diese Weiterleitung basiert auf den Routen, die in der Datei `conf/routes` definiert sind.

```
GET    /projects/:id    ProjectController.read(id: String)
```

Eine Route besteht dabei immer aus einer der HTTP-Methoden GET, POST, HEAD, PUT oder DELETE (vgl. [Pla13]), einem URI-Pattern und einer Action, die den Request beantwortet.

2. Die Action im Controller ist für die Beantwortung zuständig. Dazu können Informationen im Model abgefragt werden. Die Templates können benutzt werden, um ein dynamisches Ergebnis für den Client zu erzeugen.

SecureSocial umfasst den User-Login. Dieses Paket ist ein Authentifizierungsmodul mit Support für OAuth¹², OAuth2¹³, OpenID¹⁴ und Username/Passwort-Authentifizierung.

3.1.4 Modularisierung

Play erlaubt die Modularisierung des Codes in sogenannte Subprojekte. Ein solches Subprojekt ist dabei eine abgeschlossene Einheit, welche allein kompiliert, getestet und ausgeführt werden kann. Dabei können neben Play-Projekten auch Java- oder Scala-Projekte als Subprojekte verwendet werden. Die einzelnen Projekte und deren Abhängigkeiten werden in der *Build.scala*-Datei angegeben. Project-Zoom besteht aus drei verschiedenen Subprojekten:

common enthält den Quelltext, der sowohl von den Projekten *main* als auch *admin* benötigt wird. In diesem Projekt sind die Datenmodelle definiert. Weiterhin befinden sich in diesem Modul das Eventsystem, die Erweiterungen zum Datenaggregieren und zum Thumbnailgenerieren, sowie die Authentifizierung.

¹²OAuth, <https://tools.ietf.org/html/rfc5849> (Zugriff 20.06.13)

¹³OAuth2 <https://tools.ietf.org/html/rfc6749> (Zugriff 20.06.13)

¹⁴OpenID, http://openid.net/specs/openid-authentication-2_0.html (Zugriff 20.06.13)

main schließt alle Controller ein, die für den normalen Nutzer ansprechbar sind. Es werden verschiedene Actions definiert und für die jeweiligen Sichten Templates angelegt. Der Frontend-Code, welcher näher in den Arbeiten [Rze13], [Her13] und [Die13] beschrieben ist, wird ebenfalls in diesem Projekt verwaltet.

admin definiert jede Interaktion, die nur für privilegierte Nutzer sichtbar sein soll. Dies sorgt für eine klare Trennung zwischen User- und Admin-Anfragen und gewährleistet mehr Sicherheit (vgl. Anforderung F1). Ein weiterer Vorteil ist, dass durch das Abschalten dieses Subprojektes jedwede Admin-Aktion unterbunden werden kann.

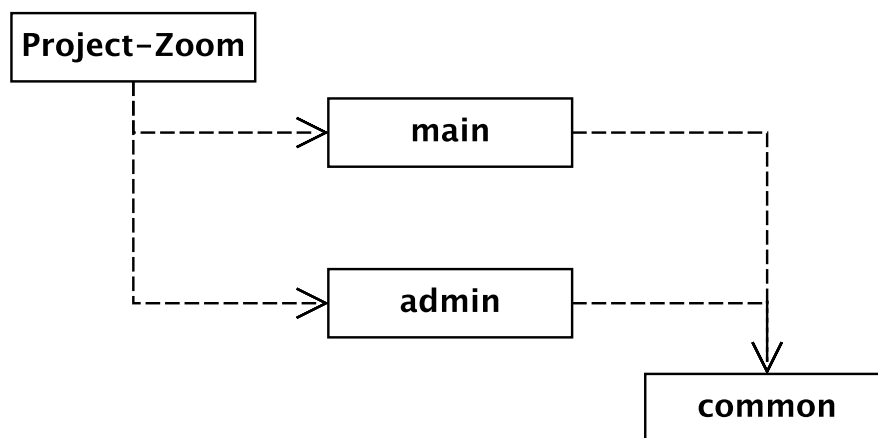


Abbildung 3.1 – Projekte und deren Abhängigkeiten

In der Abbildung 3.1 sind die Abhängigkeiten der Pakete voneinander dargestellt. Das Anlegen eines Hauptprojektes, in diesem Fall **Project-Zoom**, erleichtert die Arbeit mit dem Gesamtprojekt. Durch die Abhängigkeit zu *main* und *admin* muss nur noch das Hauptprojekt kompiliert werden, denn die Abhängigkeiten werden bei Änderungen im Source-Code automatisch mit kompiliert.

3.2 Data-centric Design in Project-Zoom

Die REST-Schnittstelle von **Project-Zoom** entspricht der einer klassischen CRUD-Anwendung. Eine Übersicht über alle REST-Routen finden sich im Anhang D. Aufgrund der geringen Business-Logik liegt es nahe, das beschriebene data-centric Design zu verwenden. Um die Wartung des Systems zu erleichtern, werden Datenzugriffsobjekte für jedes Datenmodell verwendet.

Quelltext 3.1 – Macro-Verwendung zur Generierung einer JSON-Konvertierung

```
case class User(firstName: String, lastName: String)

val userFormat = Json.format[User]
```

Für einige Teile der Anwendung ist es notwendig die JSON-Strukturen der Datenbank in Scala-Objekte umzuwandeln. Dies ist zum Beispiel für User-Objekte der Fall, welche von der Authentifizierungsbibliothek SecureSocial verwendet werden. Play erlaubt die Generierung der Transformationsfunktionen mit Hilfe von Macros¹⁵. Das bedeutet, dass die Anwendung die Datenklassen des Modells definiert und die Funktionen zur Konvertierung von und nach JSON während der Kompilierung vom Framework übernommen werden. Ein Beispiel für solch eine Transformation ist in dem Quelltext 3.1 mit der Variablen *userFormat* gegeben. Diese ist in der Lage, zwischen den beispielhaften Repräsentationen

```
{
  "firstName" : "Max",
  "lastName"  : "Mustermann"
}
```

und

```
User("Max", "Mustermann")
```

zu konvertieren. Die Bedarfskonvertierung ist durch die Verwendung von unveränderbaren Case-Klassen typsicher und entspricht den Konzepten der funktionalen Programmierung (vgl. [For11]).

Die CRUD-Operationen *list* und *read* wurden implementiert, ohne Datenbankobjekte zu konvertieren. Die Daten werden aus der Datenbank gelesen, transformiert und anschließend an den anfragenden Client zurückgegeben. Bei der Transformation werden z.B. Passwort-Hashes und Login-Informationen, die sich im User-Objekt befinden, entfernt. Die Funktionen *create*, *update* und *delete* sind nicht für alle Datenbankobjekte implementiert. Dies liegt an der Anforderung, dass die Datenhaltung in der von der D-School bereits verwendeten Datenbank geschehen soll. Project-Zoom soll diese Daten nur aggregieren.

3.3 Anbindung externer Komponenten

Die Implementierung des Eventsystems zur Anbindung externer Komponenten in Project-Zoom basiert auf dem Akka-Framework. Die in Absatz 3.1.3 beschriebenen Akteure eignen sich sehr

¹⁵JSON Macro inception, <http://www.playframework.com/documentation/2.1.1/ScalaJsonInception> (Zugriff 26.06.13)

gut zur Modellierung eines Subscribers. Ein Akteur ist eine abgeschlossene funktionelle Einheit, deren Kommunikation nur über unveränderbare Nachrichten erfolgt. Erhält ein Akteur eine Nachricht, so wird diese in der Mailbox gesammelt. Die Nachrichten in der Mailbox werden dann mit Hilfe der `RECEIVE`-Funktion, welche jeder Akteur definieren muss, abgearbeitet.

In der Umsetzung erfolgt das Abonnieren einer Nachricht durch das Senden einer partiellen Funktion¹⁶ an den Dispatcher. Diese Funktion bildet den Typ *Event* auf *Unit*¹⁷ ab und ist auf allen Events definiert, welche der Subscriber empfangen will. Der Dispatcher benutzt anschließend alle partiellen Funktionen, um eine eingehende Nachricht zu verteilen. Standardmäßig hat ein Subscriber alle Nachrichten abonniert, für die seine `RECEIVE`-Funktion definiert ist.

3.3.1 Anbindung der Konnektoren und des Thumbnailgenerators

Externe Komponenten kommunizieren untereinander und mit dem Core über dieses beschriebene Eventsystem. Ein Beispiel solch einer Kommunikation ist in der Abbildung 3.2 zu sehen. Das Sequenzdiagramm zeigt wie ein Konnektor ein neues Artefakt registriert, welches anschließend vom Core und vom Thumbnailsystem verarbeitet wird. Eine Übersicht über alle vom Backend-Core empfangenen und gesendeten Events findet sich im Anhang C.5.

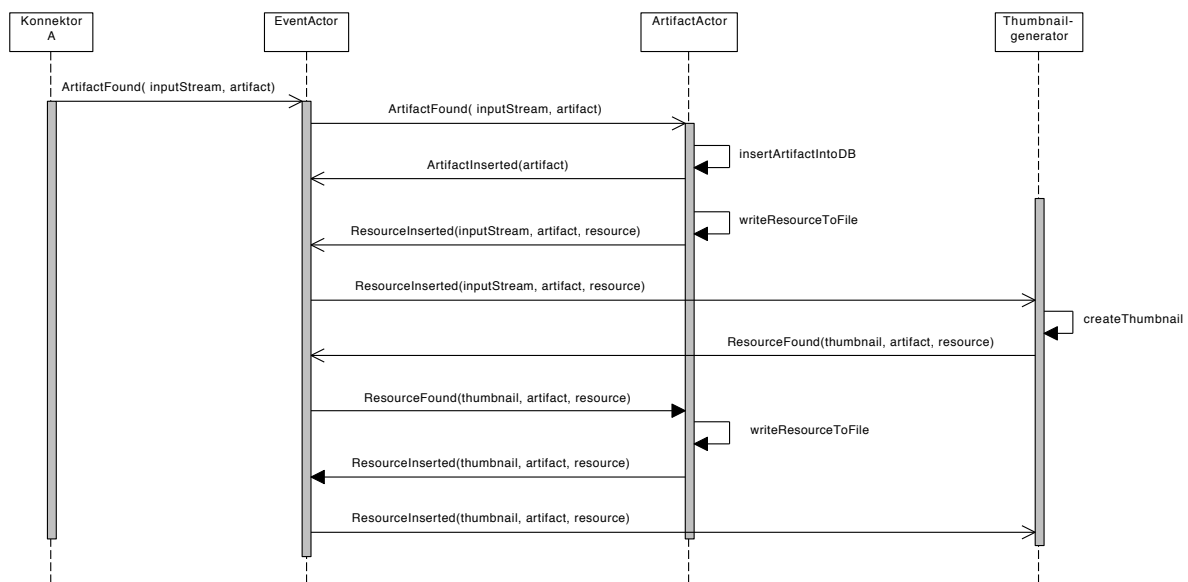


Abbildung 3.2 – Beispielhafter Ablauf beim Fund eines neuen Artefakts

¹⁶Eine partielle Funktion $f : A \rightarrow B$ ist im Gegensatz zu einer totalen Funktion nicht auf jedem Wert aus A definiert. Für die Anwendung in Scala siehe [Bru11].

¹⁷Der Rückgabetyt Unit entspricht dem Java Pendant *void*.

3.4 Evaluierung der Anforderungen

Den Abschluss der Betrachtungen zur Implementierung soll ein Rückblick auf die im Abschnitt 1.1 aufgestellten Anforderungen an das System bilden.

Es wurde ein System entwickelt, welches zur Trennung von Administration und Nutzung in Module geteilt ist. Die Webanwendung ist durch das Authentifizierungsmodul SecureSocial und die implementierten Autorisierungsmechanismen in den Datenbankzugriffsobjekten geschützt (vgl. Anforderung F1). Die Anwendung ist aus dem Internet erreichbar und somit auch außerhalb der D-School verfügbar (vgl. Anforderung F6).

Bei der Umsetzung der Konnektoren (vgl. Anforderung F2) sei auf die Arbeit von Thomas Werkmeister [Wer13], in welcher die umgesetzten Konnektoren erläutert sind, verwiesen. Die Datenquellen die in [Wer13] beschrieben sind, werden durch die Aggregation nicht verändert (vgl. Anforderung F5). Die Anforderung F3 verlangt, dass keine Daten gelöscht werden. Dies wurde umgangen, indem beim Löschen von Elementen aus einer aggregierten Datenquelle die entsprechenden Artefakte als gelöscht markiert werden.

Die Anforderung F4 stellt den Anspruch auf eine einfache Anbindbarkeit von neuen Datenquellen. Durch das Eventsystem, welches durch den Backend-Core zur Verfügung gestellt wird, entsteht ein System, welches viel Spielraum für das Anbinden neuer Konnektoren bietet.

Um die maximale Last des Systems zu prüfen, wurde mit Hilfe von *Apache HTTP Bench*¹⁸ ein Lasttest ausgeführt. Bei diesem Test betrug die ermittelte, maximale Last im Durchschnitt 461 Anfragen pro Sekunde (vgl. Anforderung NF1). Nähere Informationen und Ergebnisse finden sich im Anhang C.4.

Die Aktualisierung der angebundenen Datenquellen erfolgt einmal pro Minute (vgl. [Wer13]). Damit ist auch die Anforderung NF2 erfüllt.

¹⁸Apache HTTP Server Benchmarking Tool, <http://httpd.apache.org/docs/2.2/programs/ab.html> (Zugriff 29.06.13)

4 Fazit und Ausblick

In dieser Arbeit wurde der konzeptionelle Aufbau des umgesetzten IT-Systems für die HPI School of Design Thinking näher erläutert. Es wurde die Architektur, beginnend beim Datenmodell, über das Konzept des data-centric Designs, bis hin zur Anbindung von externen Datenquellen dargelegt. Dabei wurden die Anforderungen des Projektpartners einbezogen und deren Umsetzung am Ende evaluiert.

Während der Evaluierung des Systems unterbreiteten die Testenden verschiedene Verbesserungen an der Nutzeroberfläche (vgl. [Her13]). Diese gilt es zu analysieren und mit den Anforderungen der D-School abzugleichen.

Im Zuge der Umsetzung des Projektes sind verschiedene Ideen aufgekommen, das System zu erweitern. Ein Beispiel ist der Mehrbenutzerbetrieb, in welchem mehrere Studenten gleichzeitig an unterschiedlichen Rechnern den Graphen eines Projektes ändern können. Dem Backend fällt hier die Aufgabe zu, diese Änderungen zusammenzuführen und die einzelnen Nutzer über den neusten Stand zu informieren.

Die Administration des Systems ist ein weiterer Punkt, dem bei der zukünftigen Weiterentwicklung des Projektes mehr Aufmerksamkeit zukommen sollte. Funktionalitäten, welche für das erste Testen unwesentlich waren, wie das Administrieren der Nutzer und Projekte, müssen für den Produktivbetrieb umgesetzt werden. Dabei muss in Zusammenarbeit mit der D-School ein Konzept des Datenaustausches mit der internen Projektverwaltung entwickelt und anschließend die entsprechenden Funktionalitäten im Backend umgesetzt werden.

Der Erfolg und die Akzeptanz des Projektes als Dokumentationsvisualisierer hängt vorrangig von der zukünftigen Entwicklung der Nutzerschnittstelle ab. Das Konzept des Systems wurde von den Studenten sehr gut aufgenommen. Das Ziel muss es also sein, durch die Fortführung der Analyse der Nutzeranforderungen, den kreativen Ideenaustausch mit den Mitarbeitern und Studierenden der D-School aufrechtzuerhalten, um dadurch die Dokumentation der D-School mit Project-Zoom weiter zu optimieren.

Literaturverzeichnis

- [ACM03] Deepak Alur, John Crupi, and Dan Malks. *Core J2EE Patterns: Best Practices and Design Strategies*. Prentice Hall / Sun Microsystems Press, USA, California, June 2003.
- [Akk12] Akka. What is Akka? Webseite, 2012. Erreichbar unter: <http://doc.akka.io/docs/akka/2.0/intro/what-is-akka.html>; besucht am 19. Juni 2013.
- [Amb03] Scott W. Ambler. Mapping Objects to Relational Databases. Webseite, 2003. Erreichbar unter: <http://www.agiledata.org/essays/mappingObjects.html>; besucht am 15. Mai 2013.
- [Atw06] Jeff Atwood. Object-Relational Mapping is the Vietnam of Computer Science. Webseite, 2006. Erreichbar unter: <http://www.codinghorror.com/blog/2006/06/object-relational-mapping-is-the-vietnam-of-computer-science.html>; besucht am 10. Juni 2013.
- [BBD⁺13] Dominic Bräunlein, Tom Bocklisch, Anita Diekhoff, Norman Rzepka, Tom Herold, and Thomas Werkmeister. Software Requirements Specification. 2013.
- [BH77] Henry C. Baker, Jr. and Carl Hewitt. The incremental garbage collection of processes. In *Proceedings of the 1977 symposium on Artificial intelligence and programming languages*, pages 55–59, New York, NY, USA, 1977. ACM.
- [Bow10] David Bowers. Object Relational Database Design – impedance mismatch or expectation mismatch? Webseite, 2010. Erreichbar unter: <http://www.dmsg.bcs.org/web/images/stories/PDFs/2010-10-13-david-bowers.pdf>; besucht am 23. Juni 2013.
- [Brä13] Dominic Bräunlein. Generierung und Bereitstellung von semantischen Thumbnails aus heterogenen Daten für Project-Zoom, 2013.
- [Bru11] Erik Bruchez. Scala partial functions (without a PhD). Webseite, 2011. Erreichbar unter: <http://blog.bruchez.name/2011/10/scala-partial-functions-without-phd.html>; besucht am 10. Juni 2013.
- [CLRS01] Th. H. Corman, Ch. E. Leiserson, R. Rivest, and C. Stein. *Algorithmen - Eine Einführung*. MIT Press, 2001.

- [Die13] Anita Diekhoff. Automatisches Layouten in interaktiven Graphen, 2013.
- [Dir10] Mike Dirolf. BSON. Webseite, 2010. Erreichbar unter: <http://bsonspec.org>; besucht am 19. Juni 2013.
- [FFBS04] Eric Freeman, Elisabeth Freeman, Bert Bates, and Kathy Sierra. *Head First Design Patterns*. O'Reilly, Sebastopol, CA, October 2004.
- [Fie00] Roy Thomas Fielding. *Architectural Styles and the Design of Network-based Software Architectures*. PhD thesis, UNIVERSITY OF CALIFORNIA, 2000.
- [For11] Neal Ford. Functional thinking: Immutability - Make Java code more functional by changing less. Technical report, IBM Corporation, July 2011.
- [GHJV95] Erich Gamma, Richard Helm, Ralph Johnson, and John Vlissides. *Design patterns: elements of reusable object-oriented software*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1995.
- [HBS73] Carl Hewitt, Peter Bishop, and Richard Steiger. A universal modular actor formalism for artificial intelligence. In *Proceedings of the 3rd international joint conference on Artificial intelligence, IJCAI'73*, pages 235–245, San Francisco, CA, USA, 1973. Morgan Kaufmann Publishers Inc.
- [Her13] Tom Herold. Kontextsensitiver Assistent für interaktive Graphen für Project-Zoom, 2013.
- [HT99] Andrew Hunt and David Thomas. *The pragmatic programmer: from journeyman to master*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1999.
- [HVE03] Anders Hejlsberg, Bill Venners, and Bruce Eckel. Inappropriate Abstractions - A Conversation with Anders Hejlsberg, Part VI. Webseite, 2003. Erreichbar unter: <http://www.artima.com/intv/abstract3.html>; besucht am 10. Juni 2013.
- [IBNW09] Christopher Ireland, David Bowers, Michael Newton, and Kevin Waugh. A classification of object-relational impedance mismatch. In *Advances in Databases, Knowledge, and Data Applications*, pages 36–43. IEEE, 2009.
- [LM10] Adam Lith and Jakob Mattsson. Investigating storage solutions for large data. Master's thesis, Chalmers University Of Technology, 2010.
- [New08] Ted Neward. *The Busy Java Developer's Guide to Scala: Of Traits and Behaviors. IBM developerWorks*, 2008.
- [Ode11] Martin Odersky. *Scala By Example*. EPFL, Switzerland, May 2011.
- [PGB⁺05] Tim Peierls, Brian Goetz, Joshua Bloch, Joseph Bowbeer, Doug Lea, and David Holmes. *Java Concurrency in Practice*. Addison-Wesley Professional, 2005.

- [Pla13] Play. HTTP routing. Webseite, 2013. Erreichbar unter: <http://www.playframework.com/documentation/2.1.1/ScalaRouting>; besucht am 19. Juni 2013.
- [PMW09] H. Plattner, C. Meinel, and U. Weinberg. *Design Thinking: Innovation lernen, Ideenwelten öffnen*. Mi-Wirtschaftsbuch, FinanzBuch-Verlag, 2009.
- [Pok11] Jaroslav Pokorny. Nosql databases: a step to database scalability in web environment. In *Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services*, iiWAS '11, pages 278–283, New York, NY, USA, 2011. ACM.
- [Ree00] George Reese. *Database Programming with JDBC and JAVA*. O'Reilly Media, Inc., 2000.
- [Rze13] Norman Rzepka. The design of an web-based event-driven client application architecture for Project-Zoom, 2013.
- [SA12] Aaron Schram and Kenneth M. Anderson. Mysql to nosql: data modeling challenges in supporting scalability. In *Proceedings of the 3rd annual conference on Systems, programming, and applications: software for humanity*, SPLASH '12, pages 191–202, New York, NY, USA, 2012. ACM.
- [Sca08] Scala. A Tour of Scala: Case Classes. Webseite, 2008. Erreichbar unter: <http://www.scala-lang.org/node/107>; besucht am 17. Juni 2013.
- [Voi13] Pascal Voitet. JSON Coast-To-Coast. Webseite, 2013. Erreichbar unter: <http://mandubian.com/2013/01/13/JSON-Coast-to-Coast/>; besucht am 21. März 2013.
- [Wer13] Thomas Werkmeister. Extensible backend architecture for data aggregation of heterogeneous sources for Project-Zoom, 2013.
- [WP10] D. Wampler and A. Payne. *Programmieren mit Scala*. O'Reilly Vlg. GmbH & Company, 2010.
- [Zyl08] Jason Van Zyl. *Maven - the definitive guide: everything you need to know from ideation to deployment*. O'Reilly, 2008.

Glossar

Artefakt Eine Informationseinheit, welche sich in dieser Arbeit auf die beim Dokumentationsprozess entstehenden, digitalen Informationen bezieht. 4, 9–11, 18, 22, 28, 29, III

Asynchronität Beschreibt die zeitliche Unabhängigkeit von zwei Ereignissen. 21

Backend Schicht einer Anwendung, welche bei einer Client-Server Applikation den Anwendungscode des Servers bezeichnet. 3–5, 7, 8, 10, 12–14, 21, 24, 25, 28, 29, 31, III

Business-Logik Teil eines Systems, welcher die Logik für die Umsetzung der Problemstellung implementiert. 7, 14, 15, 18, 26

Eventsystem Ein System, dessen Architektur auf dem Zusammenspiel der Komponenten durch *Events* (Ereignissen) basiert. 5, 7, 13, 18, 19, 21, 25, 27–29, III

Framework In der Softwareentwicklung ein Rahmen, der unterschiedlichste Funktionalitäten zur Verfügung stellt und innerhalb dessen der Anwendungscode ausgeführt wird. 8, 11, 15, 21, 22, 24, 25, 27

Frontend Schicht einer Anwendung, welche bei einer Client-Server Applikation den Anwendungscode des Client bezeichnet. 5, 7, 10, 12, 25

Konnektor Komponente, die den Zugriff auf externe Dienste ermöglicht. 5, 28, 29

Netzwerkdateisystem Ein Dateisystem, auf das von Rechnern im selben Netzwerk zugegriffen werden kann, um zum Beispiel Dateien zu lesen. 2, 18

Ressource Die digitale Repräsentation eines Artefakts. Verschiedene Thumbnails eines Artefakts sind jeweils eine Ressource. 2, 8, 11, 14, 23

Thumbnail Miniaturbild, welches als Vorschau für ein Bild oder eine Datei dient. 11, 18, 19, 21, 25, 28, 37

Unveränderbarkeit Bezüglich des Statuses eines Objektes bedeutet dies, dass zur Laufzeit Attributen eines Objektes nach Erstellung dessen keine neuen Werte mehr zugewiesen werden können (vgl. [PGB⁺05, p. 31]). 16

Anhang A

Quelltexte

A.1 Quelltext-Beispiel für eine JSON-Transformation und -Validierung

Im Quelltext A.1 ist die Transformierung und Validierung eines JSON-Objektes gezeigt. Zeilen, welche mit `//>` beginnen, sind Kommentare und zeigen das Ergebnis der vorhergehenden Anweisung.

Quelltext A.1 – Beispiel für Play JSON-Transformation bzw. -Validierung

```
val json = Json.obj(
  "firstName" -> "max",
  "lastName" -> "Mustermann",
  "age" -> 17)

val addUserRole = JsPath.json.update(
  (JsPath \ "role").json.put(JsonString("User")))
//> addUserRole : Reads[JsonObject]

val isMajor =
  (JsPath \ "age").read[Int](min(18))
//> isMajor : Reads[Int]

json.transform(addUserRole)
//> res0: JsResult[JsonObject] =
//    JsSuccess({
//      "firstName": "max",
//      "lastName": "Mustermann",
//      "age": 17,
//      "role": "User"},)

json.validate(isMajor)
//> res1: JsResult[Int] =
//    JsError(List((
//      /age,List(
//        ValidationErrors(validate.error.min,WrappedArray(18))))))
```


Anhang B

Interview Protokolle

B.1 Interview 22.11.2012 - Claudia Nikolai

Claudia Nikolai ist General Programm Manager der D-School.

Zuständigkeiten

- inhaltliche Programmgestaltung
- internationale Kooperation
- Kontakt zu Kunden
- Ausbildung der Teacher
- Teacher (an allen Tagen)

Austausch mit anderen Teachern

- Teachermeeting
- E-Mail (Verteiler, Informationen fuer viele Personen)
- Dropbox/E-mail für Content

Incom

- space noch nicht privat/sichtbar (kein Austausch von vertraulichen Informationen)
- zu viele Notifications (liest aber alle)
- nutzt es um Hinweise an andere Teams zu geben
- sichtet Material an Tagen, an denen keine D-School ist

Wichtig für die Dokumentation

- Wie haben die Studenten Empathie bekommen? (Wie und wann haben sie sich in den User hineinversetzt?)
- Zitate
- Dokumentation der Phasen
- Welche Methode wurde verwendet und wie gut hat sie funktioniert? -> Evaluierung des DS-Prozesses

Wunschsystem

- System sollte einfach verständlich sein, Nutzer haben nur 6–12 Wochen um damit klar zu kommen
- System muss keine eierlegende Wollmilchsau sein

Anhang C

Ergänzende Visualisierungen

C.1 Datenmodell von Project-Zoom

In der Abbildung C.1 ist das komplette Datenmodell von Project-Zoom dargestellt.

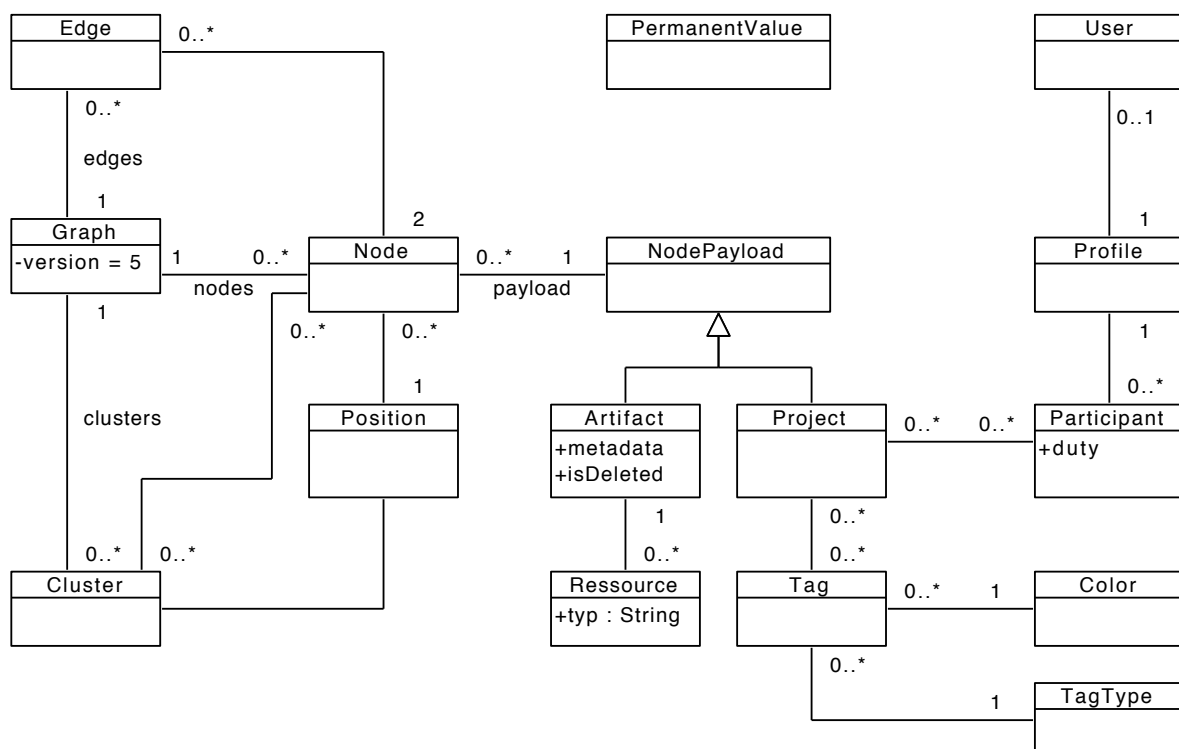


Abbildung C.1 – Datenmodell von Project-Zoom

C.2 Objektdiagramm für einen beispielhaften Graphen

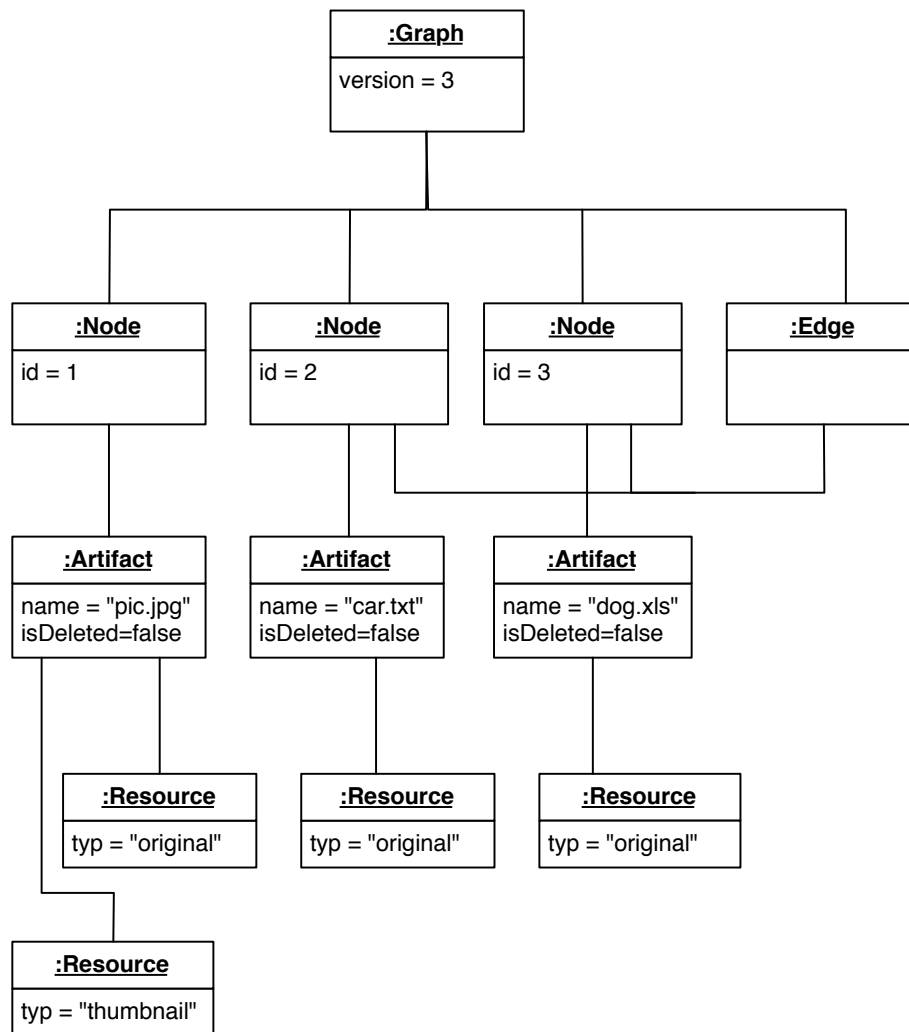


Abbildung C.2 – Objektdiagramm eines beispielhaften Graphen. Zur Wahrung der Übersichtlichkeit ist die Position jedes Nodes nicht als Objekt visualisiert.

C.3 InfoQ Umfrage zu Webframeworks der JVM

In der Tabelle in der Abbildung C.3 sind die Umfrageergebnisse der InfoQ-Umfrage mit 1894 Teilnehmern von der Seite <http://www.infoq.com/research/jvm-web-frameworks.auf> zum angegebenen Zeitpunkt zu sehen.

Total number of participants: **1894**






















Option	Adoption Ready	Importance	Votes	Details
Spring MVC	85%	82%	1056	Heatmap 
Play	71%	78%	875	Heatmap 
Grails	77%	76%	833	Heatmap 
JSF	75%	65%	712	Heatmap 
Struts	80%	62%	676	Heatmap 
Wicket	72%	70%	478	Heatmap 
Lift	61%	65%	425	Heatmap 
Seam	68%	64%	386	Heatmap 
Tapestry	72%	67%	382	Heatmap 
JRoR	66%	65%	320	Heatmap 
Vert.x	51%	68%	311	Heatmap 
Netty	69%	74%	310	Heatmap 
Sitemesh	70%	64%	292	Heatmap 
Scalatra	56%	63%	268	Heatmap 
Stripes	62%	60%	226	Heatmap 
Compojure	54%	55%	214	Heatmap 
Conjure	53%	52%	204	Heatmap 
Cascade	56%	47%	198	Heatmap 
VRaptor	59%	67%	191	Heatmap 
Noir	52%	59%	186	Heatmap 
Spark	48%	56%	148	Heatmap 
Adopt	Trial	Assess	Hype	

Abbildung C.3 – Stand der Umfrage am 26. Juni 2013

C.4 Performance Test

Project-Zoom wurde mit den folgenden Einstellungen für Apache Bench getestet:

```
ab -g output100.txt -c 10 -n 100000 http://localhost:9000/projects
```

Eigenschaften des getesteten Systems:

- 4 Kerne + HT
- 8GB RAM
- 256GB SSD Festplatte
- OS X 10.8.4
- ApacheBench, Version 2.3 Revision: 1430300
- Java Version 1.7.0_21 / Scala Version 2.10.1 / Play Version 2.1

Zuzüglich zu den angegebenen Parametern wurde noch ein Session-Cookie übergeben, um auf die geschützte Ressource zugreifen zu dürfen. Die Ergebnisse sind in den Abbildungen C.4 und C.5 zu sehen. Die Grafiken wurden mit Hilfe von <https://loadosophia.org> erstellt.

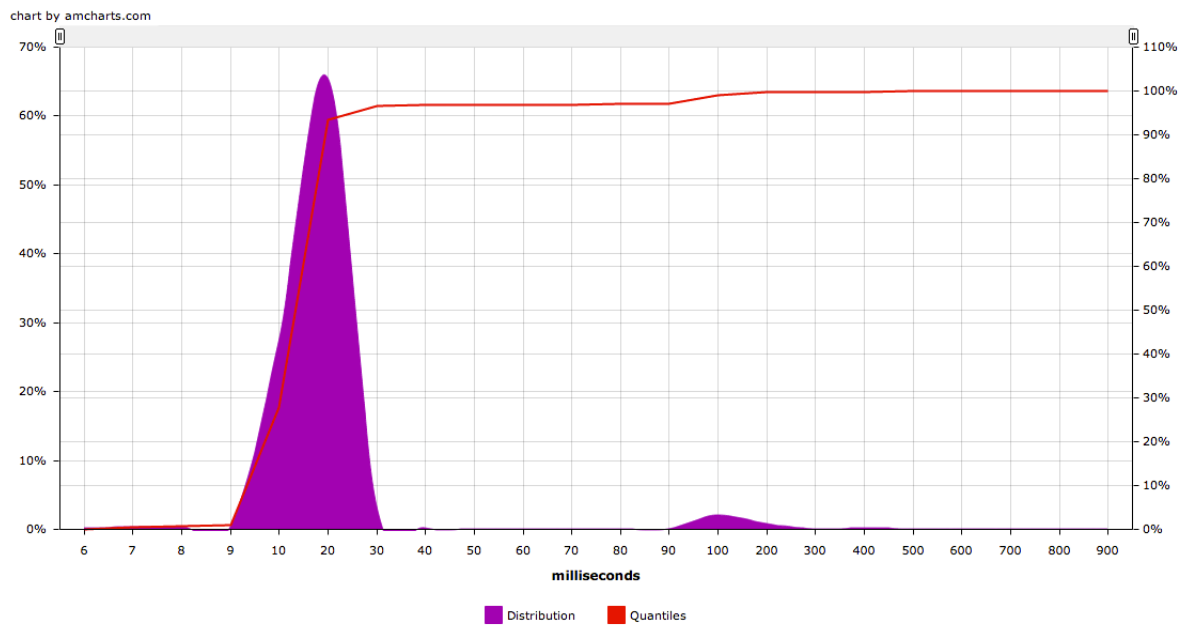


Abbildung C.4 – Dichte- und Verteilungsfunktion der Antwortzeiten

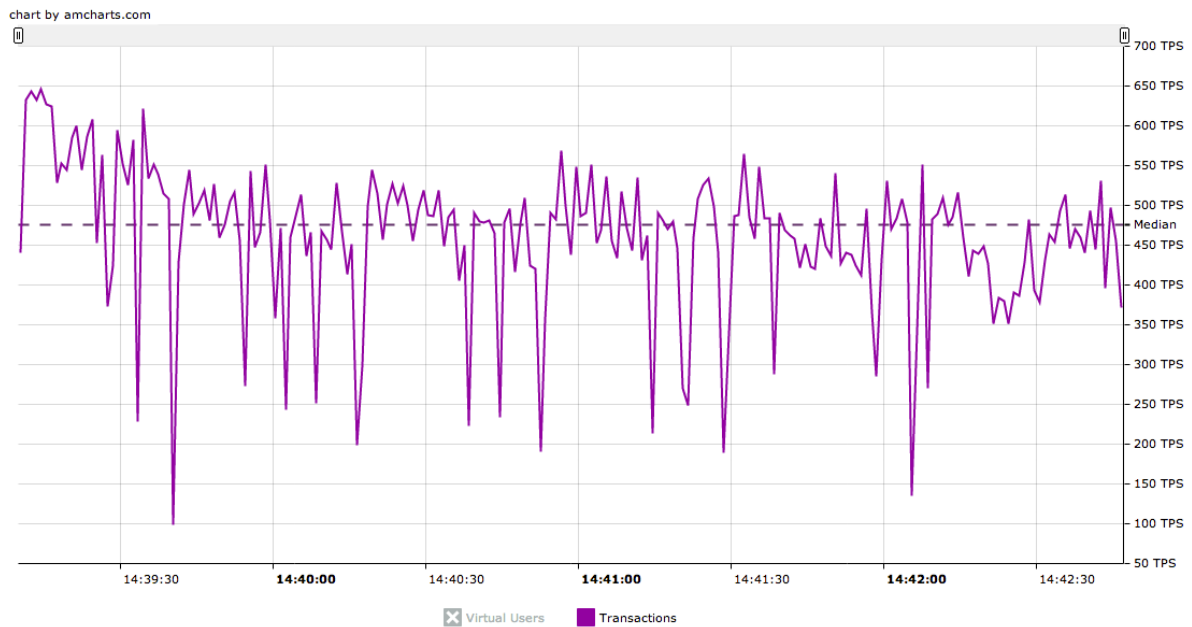


Abbildung C.5 – Durchsatz in abgeschlossenen Übertragungen pro Sekunde über die Zeit

C.5 Überblick über die vorhandenen Events

In der Tabelle C.1 sind alle Events zu sehen, die mit dem Backend-Core interagieren.

ARTIFACTACTOR

Publishes: ArtifactUpdated, ArtifactInserted, ResourceUpdated, ResourceInserted

Subscription für: ArtifactFound, ArtifactDeleted, ArtifactAggregation, ArtifactRenamed, ArtifactMoved, ResourceFound

KNOWLEDGEACTOR

Subscription für: ProjectFound, ProjectAggregation, ProfileFound, ProfileAggregation

KONNEKTOREN

Publishes: ProfileAggregation, ProjectAggregation, ArtifactFound, ArtifactRenamed, ArtifactMoved, ArtifactDeleted, ArtifactAggregation

THUMBNAILGENERATOR

Publishes: ResourceFound

Subscription für: ResourceInserted, ResourceUpdated

	Parameter	Bedeutung
ArtifactFound	<code>originalStream: InputStream</code> <code>artifact: ArtifactLike</code>	Artefakt von Konnektor gefunden
ArtifactDeleted	<code>artifact: ArtifactLike</code>	Löschung von Artefakt durch Konnektor festgestellt
ArtifactAggregation	<code>_project: String</code> <code>l: List[ArtifactFound]</code>	Auflistung aller Artefakte eines Projektes durch Konnektor
ArtifactRenamed	<code>artifact: ArtifactLike</code> <code>name: String</code>	Umbenennung von Artefakt durch Konnektor festgestellt
ArtifactMoved	<code>artifact: ArtifactLike</code> <code>path: String</code>	Verschiebung von Artefakt durch Konnektor festgestellt
ResourceFound	<code>inputStream: InputStream</code> <code>artifact: ArtifactLike</code> <code>resource: ResourceLike</code>	Neue Resource für Artefakt gefunden
ArtifactUpdated	<code>artifact: ArtifactLike</code>	Artefakt in der DB geändert
ArtifactInserted	<code>artifact: ArtifactLike</code>	Artefakt zur DB hinzugefügt
ResourceUpdated	<code>file: File</code> <code>artifact: ArtifactLike</code> <code>resource: ResourceLike</code>	Resource in der DB geändert
ResourceInserted	<code>file: File</code> <code>artifact: ArtifactLike</code> <code>resource: ResourceLike</code>	Resource zur DB hinzugefügt
ProjectFound	<code>project: ProjectLike</code>	Projekt von Konnektor gefunden
ProjectAggregation	<code>l: List[ProjectFound]</code>	Auflistung mehrerer gefundener Projekte
ProfileFound	<code>profile: Profile</code>	Profil von Konnektor gefunden
ProfileAggregation	<code>l: List[ProfileFound]</code>	Auflistung mehrerer gefundener Profile
GraphUpdated	<code>graph: Graph</code> <code>patch: JsValue</code>	Graph in der DB mit Patch geändert

Tabelle C.1 – Übersicht über Events des Backend-Core

Anhang D

REST-Schnittstelle

Datentypen

identifier	example
int	3
double	3.5
bool	true
string	"test"
object_id	"512d2218c2c1804377000005" ¹
"date"	"2013-02-26T02:00:00Z" ²
{object}	{ "test": 123 }

Allgemein

`GET /resources`

returns the 50 first resources

`GET /resources?limit=100`

returns the 100 first resources

`GET /resources?offset=30&limit=10`

returns 10 resources from the 30th

`GET /resources/:id`

returns all details of a specific resource

¹Mongo ObjectId, <http://docs.mongodb.org/manual/reference/object-id/>, accessed 25/06/13

²ISO 8601, http://en.wikipedia.org/wiki/ISO_8601, accessed 25/06/13

```
PUT /resources/:id
# replaces this item
=> 200 OK
or
=> 400 Bad Request
{
  errors: []
}
```

```
PATCH /resources/:id
# incrementally updates this item
=> 200 OK
or
=> 400 Bad Request
{
  errors: []
}
```

```
POST /resources
# creates a new item
=> 200 OK
or
=> 400 Bad Request
{
  errors: []
}
```

```
DELETE /resources/:id
# deletes this item
=> 200 OK
or
=> 400 Bad Request
{
  errors: []
}
```


Projekte

GET /projects

```
{
  "limit": int,
  "offset": int,
  "content": [
    {
      "id": "object_id",
      "name": "string",
      "length": "string",
      "season": "string",
      "year": "string",
      "_graphs": ["object_id", ...],
      "_tags": ["object_id", ...]
    }, ...
  ]
}
```

GET /projects/:id

```
{
  "id": "object_id",
  "name": "string",
  "length": "string",
  "season": "string",
  "year": "string",
  "_graphs": ["object_id", ...],
  "_tags": ["object_id", ...]
}
```

GET /projects/:id/artifacts

```
{
  "limit": int,
  "offset": int,
  "content": [
    {
      "id": "object_id",
      "name": "string",
      "path": "string",
      "projectName": "string",
      "resources": [
        {
          "hash": "string",

```

```
        "name": "string",
        "typ": "string"
      }, ...
    ],
    "createdAt": int,
    "isDeleted": bool,
    metadata: {object}
  }, ...
]
```

```
POST /projects/:id/graphs
=> 303 See other
Location: /graphs/:graph_group
```

Artefakte

```
GET /artifacts/:id
{
  "id": "object_id",
  "name": "string",
  "path": "string",
  "projectName": "string",
  "resources": [
    {
      "hash": "string",
      "name": "string",
      "typ": "string"
    }, ...
  ],
  "createdAt": int,
  "isDeleted": bool,
  metadata: {object}
}
```

```
GET /artifacts/:id/:typ/:name
=> 200 OK
Content-Type: application/octet-stream
```

Tags

GET /tags

```
{
  "limit": int,
  "offset": int,
  "content": [
    {
      "color": {"r": int, "g": int, "b": int},
      "id": "object_id",
      "name": "string"
    }, ...
  ]
}
```

GET /tags/:name

```
{
  "color": {"r": int, "g": int, "b": int},
  "id": "object_id",
  "name": "string"
}
```

User

GET /users

```
{
  "limit": int,
  "offset": int,
  "content": [
    {
      "id": "object_id",
      "email": "string"
      "firstName": "string",
      "lastName": "string"
    }, ...
  ]
}
```

```
GET /users/:email
{
  "id": "object_id",
  "email": "string"
  "firstName": "string",
  "lastName": "string"
}
```

Graphen

```
GET /graphs
{
  "limit": int,
  "offset": int,
  "content": [
    {
      "_project": "object_id",
      "id": "object_id",
      "clusters": [
        {
          "comment": "string",
          "content": [int, int, ...],
          "id": int,
          "phase": "string",
          "waypoints": [{ "x": int, "y": int }, ...]
        }, ...
      ],
      "edges": [
        {
          "from": int,
          "to": int
        }, ...
      ],
      "group": "string",
      "nodes": [
        {
          "id": int,
          "typ": "string",
          "position": { "x": int, "y": int },
          "payload": {object}
        }, ...
      ],
      "version": int,
```

```
    }, ...
  ]
}

GET /graphs/:graph_group
{
  "_project": "object_id",
  "id": "object_id",
  "clusters": [
    {
      "comment": "string",
      "content": [int, int, ...],
      "id": int,
      "phase": "string",
      "waypoints": [{ "x": int, "y": int }, ...]
    }, ...
  ],
  "edges": [
    {
      "from": int,
      "to": int
    }, ...
  ],
  "group": "string",
  "nodes": [
    {
      "id": int,
      "typ": "string",
      "position": { "x": int, "y": int },
      "payload": {object}
    }, ...
  ],
  "version": int,
}

GET /graphs/:graph_group/:version
# Same as GET /graphs/:graph_group

PATCH /graphs/:groupId/:version
{
  "version": int
}
```

Eidesstattliche Erklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbstständig verfasst und dafür keine anderen als die genannten Quellen und Hilfsmittel verwendet habe.

Tom Bocklisch

Potsdam 01. July 2013
