

## EIS WS1516 - Meilenstein 2

Verteiltes Training einer automatisierten Dokumentenattributierung

Tim Howe

TaARs - Verteilte Group- Middleware zum Training einer automatisierten Attributierung von Rechnungen

# Inhaltsverzeichnis

<b>1</b>	<b>Systembeschreibung</b>	<b>4</b>
1.1	funktionale Komponenten . . . . .	4
1.2	Akteure . . . . .	4
1.3	• . . . . .	4
<b>2</b>	<b>Zielhierarchie</b>	<b>5</b>
2.1	Strategisch . . . . .	5
2.2	Taktisch . . . . .	5
2.3	Operational . . . . .	5
<b>3</b>	<b>Domänenrecherche</b>	<b>6</b>
3.1	Allgemeine Verarbeitungprozess . . . . .	6
3.2	Der Begriff der Attributierung . . . . .	6
3.3	Strukturierungsgrad . . . . .	7
3.4	Anwendungsfelder . . . . .	8
3.5	Nutzungsmerkmale . . . . .	8
<b>4</b>	<b>Marktrecherche</b>	<b>9</b>
4.1	Codia DMS . . . . .	9
4.2	InPunkto . . . . .	9
4.3	Übersicht . . . . .	10
<b>5</b>	<b>Alleinstellungsmerkmale</b>	<b>11</b>
<b>6</b>	<b>Prozess</b>	<b>12</b>
<b>7</b>	<b>Kommunikationsmodell</b>	<b>13</b>
7.1	Akteure . . . . .	13
7.2	Kommunikationsdiagramm . . . . .	14
<b>8</b>	<b>Architekturdiagramm</b>	<b>15</b>
8.1	Variante 1: API orientiert . . . . .	15
8.2	Variante 2: Nachrichten orientiert . . . . .	15
<b>9</b>	<b>Risiken</b>	<b>16</b>
9.1	Architektur . . . . .	16
<b>10</b>	<b>Proof of Concepts</b>	<b>17</b>
10.1	Definition . . . . .	17
10.1.1	Architektur . . . . .	17
10.1.2	Regel-Engine . . . . .	17
10.1.3	mobiler Client . . . . .	17
10.1.4	technisch: Steuerung Dienst . . . . .	17
10.1.5	• . . . . .	17
10.2	Durchführung . . . . .	17
<b>11</b>	<b>Methodischer Rahmen</b>	<b>18</b>

<b>12 Projektplan</b>	<b>19</b>
<b>13 Projektbegründungen</b>	<b>20</b>
13.1 Implementierung . . . . .	20
13.2 Objektbereich . . . . .	20
13.3 Nutzermodelle . . . . .	20
13.4 Vorgehensmodell . . . . .	21
<b>14 Quellen</b>	<b>22</b>

# Kapitel 1

## Systembeschreibung

Ein allgemeine Beschreibung der wichtigsten Eigenschaften des TaARs , detaillierte Beschreibungen folgen in den zugehörigen Kapiteln

### 1.1 Komponenten

#### Verwaltungsserver

1. importiert Rohdaten, erzeugen des Geschäftsobjekts
2. prüft Rohdaten auf Vollständigkeit, vorhandene Regeln, falls eine vollständige Regel
  - (a) vorliegt: Übergabe des Geschäftsobjekts an den Regel-Server
  - (b) nicht vorliegt: Übergabe des Geschäftsobjekt an einen Verwaltungsclient
3. erhält vervollständigte Geschäftsobjekte von Verwaltungs- und Fachclient
4. gibt Informationen zum Systemzustand an Steuerungsclient

#### Verwaltungsclient

1. Windows Desktop Client
2. erhält unvollständige Geschäftsobjekte vom Verwaltungsserver
3. vervollständigt diese Geschäftsobjekte
4. sendet vollständige Geschäftsobjekte an Verwaltungsserver
5. sendet unvollständige Geschäftsobjekte an Fachclient
6. verwaltet angelegte Regeln

#### Fachclient

1. Windows Desktop Client
2. funktionale Untermenge des Verwaltungsclients
3. erhält unvollständige Geschäftsobjekte vom Verwaltungsclient
4. sendet vollständige Geschäftsobjekte an Verwaltungsserver

#### Steuerungsclient

- erhält Informatinen zum Systemzustand vom Verwaltungsserver
- übergibt Priorisierungskommandos an Verwaltungsserver

#### Regel-Server

1. erhält Geschäftsobjekt von Verwaltungsserver
2. berechnet die Anwendung der Regel auf das Geschäftsobjekt
3. exportiert das Ergebnis

## **1.2 Nutzerrollen**

## **1.3 Geschäftsobjekt**

Das verarbeitete Geschäftsobjekt besteht aus den Rohdaten einer Rechnung, den zugehörigen Metadaten und einer zugewiesenen Regel für dieses Geschäftsobjekt.

# Kapitel 2

## Zielhierarchie

### 2.1 Strategisch

1. Objektbereich:  
Es muss ein Komplexitätsgrad erreicht werden der fachlich relevant und technologisch, im Rahmen des Projekts, beherrschbar ist
2. Anwendungsdomäne:  
Es soll ein Anwendungskontext mit möglichst hoher wirtschaftlicher Relevanz gefunden werden
3. Technologisch:  
Es sollen möglichst viele im beruflichen Kontext relevanten Erfahrungen... , siehe ??
4. Nutzung:  
Anwender sollen vom Ballast repetiver Aufgaben befreit werden

### 2.2 Taktisch

1. Anwendungskontext & Objektbereich:  
Es muss eine qualifizierte Entscheidungsgrundlage geschaffen werden
2. Technologisch:  
Implementierungsrelevante Entscheidungen sollen gegen den beruflichen Kontext bewertet werden
3. Nutzung:  
Es soll Automatisierungspotential identifiziert werden
4. ...

### 2.3 Operational

1. Anwendungskontext & Objektbereich:  
Es muss eine Analyse und Bewertung der in den Anwendungsdomänen genutzten Objekte, dh. Dokumentenklassen, durchgeführt werden
2. Technologisch:  
Begründen wenn vom definierten Standard abgewichen wird
3. Nutzung:
  - (a) deskriptive Aufgabenanalyse
  - (b) Automatisierungspotential identifizieren
  - (c) präskriptive Aufgabenanalyse, inkl P2
4. ...

# Kapitel 3

## Domänenrecherche

Im folgenden werden die wichtigsten Themen der Anwendungsdomäne beschrieben und einer ersten Bewertung hinsichtlich projektrelevanter Eigenschaften unterzogen.

### 3.1 Allgemeine Verarbeitungprozess

Beispielhaft die wichtigsten funktionalen Komponenten des Dokumentenverarbeitungsprozesses beschrieben, wichtig bleibt anzumerken das in konkreten Implementierungen die Funktionalitäten verschwimmen und keine klare Trennung wie hier beschrieben vorherrscht(?).

1. Extraktion:  
Manuelles oder mittels OCR automatisiertes auslesen von Informationen aus einer Dokumentendatei oder einer zugehörigen Bitmapdatei.
  - (a) optional Nacherfassung:  
Kontrolle der OCR Ergebnisse und ggf Korrektur bei unzureichender Extraktionsqualität
2. Klassifizierung:  
Beschreibt die Einordnung in eine klar abgegrenzte Menge von Dokumenttypen wie zb. Formulare, Rechnungen, Lieferscheine, Bewerbungen.
3. Attributierung:  
Beschreibt den Prozess der Zuordnung von organisations- oder fachspezifischen Attributen zu einem Dokument zur weiteren Verarbeitung innerhalb der Organisation. Die können zum Beispiel Buchungskonten, Kostenstellen, Projektnummern oder Ansprechpartner sein.
4. Export & weitere Verarbeitung:  
Übergabe der klassifizierten und attributierten Dokumente an einem betriebliches Anwendungssystem zur Buchung, Kontierung oder Archivierung.

### 3.2 Der Begriff der Attributierung

In der Praxis wird der Begriff der Attributierung für zweierlei Aspekte verwendet:

1. semantische(?) Attributierung:  
Extrahierung und Zuordnung von Metainformationen zu einem Dokument welche sich auf das Dokument, bzw die Dokumentendatei als Repräsentation des Dokuments, selbst beziehen. Dies können zb Schlagworte, Speicherort ... sein.
2. fachliche Attributierung:  
Die fachliche Attributierung ist ein Teil des Arbeitsprozesses bei welchem dem Dokument Attribute zugeordnet werden die Diese Attribute können zb kaufmännischer, steuerlicher oder juritischer Natur sein.

Für dieses Projekt wird der Begriff im Sinne der fachlichen Attributierung verwendet.

### 3.3 Strukturierungsgrad

Der Strukturierungsgrad eines Dokuments wird beschrieben durch das Ausmaß an Sicherheit mit der ein Wert eines Dokumentenattributs an einer Position im Dokument auftritt und welcher Wertebereich in diesem abgebildet wird.

1. unstrukturiert: gar keine bis geringe Positionssicherheit mit überwiegend undefiniertem Wertebereich, zb. Bewerbungsschreiben
2. semi-strukturiert: gute Positionssicherheit mit überwiegend definiertem Wertebereich, zb. Rechnungen, Lieferscheine
3. strukturiert: absolute Positionssicherheit mit klar definiertem Wertebereich, zb. genormte betriebliche oder behördliche Formulare

Mit dem Strukturierungsgrad steigt die semantische Spezialisierung sowie das Automatisierungspotential für diesen Dokumenttyp.



### 3.4 Anwendungsfelder

Tabelle 3.1: Automatisierte Dokumentenverarbeitung

Domäne	Primäre Dokumenttypen	Strukturierungsgrad	wirtschaftliche Relevanz
Buchhaltung	Rechnungen	2	++
Verwaltung	Formulare	3	+
Personal	Bewerbungen	1-2	o
Kanzleien	Schriftverkehr	1	+
Logistik	Lieferscheine	2	+
Privat	Rechnungen	2	-
	Versicherungen	1	Services für Privatanwender
	Formulare	3	nicht etabliert

### 3.5 Nutzungsmerkmale

Eine erste kurze Betrachtung der antizipierten Nutzungsmerkmale welche Entscheidungen zu Anwendungsdomäne (ref), Objektbereich (ref) und ... unterstützen soll.

Tabelle 3.2: Nutzungsmerkmale

Domäne Organisationsrolle	Arbeitsumgebung	Arbeitsgerät(?)
Buchhaltung Bürokm Fachkraft	Büro	Desktop PC
Verwaltung Bürokm oder Verwaltungs- Fachkraft	Büro	Desktop PC
Personal Personaldienstleistungskfm Fachkraft	Büro	Desktop PC
Kanzleien Bürokm Fachkraft	Büro	Desktop PC
Logistik  Bürokm Fachkraft, Lagerist	Büro ggf Mobil, in größeren Betrieben	Desktop PC ggf Tablet
Privat	Zuhause Mobil	Desktop PC Tablet, Smartphone
Privat, Organisieren von Post(?)		

# Kapitel 4

## Marktrecherche

Der Markt wird im allgemeinen von Entwicklern einzelner Komponenten und Systemhäusern bestimmt die Fremdkomponenten ggf mit Eigenentwicklungen kombinieren und so individuelle Lösungspakete schnüren welche sich immer an den in beschrieben etablierten ??Prozess der Anwendungsdomäne orientieren.

Die Komponenten lassen sich grob in die folgenden Bereiche einteilen:

- Capture: Scannen, OCR, klassifizieren,
- Workflow: Attributierung
- betriebliche Anwendungssysteme: Buchhaltung, Archivierung

Zudem sind OEM Versionen glossar durchaus gängige Praxis wodurch eine Einsicht erschwert wird.

Um dennoch einen Eindruck in die Marktsituation zu gewinnen hilft eine Betrachtung der og Bereiche in der funktional übergeordneten Ebene der DMS-Systeme, ... und deren Eigenschaften. Es werden Ansätze zweier Anbieter exemplarisch beschrieben und eine Einordnung zwischen diesen Lösungen und TaARs mittels einer Featurematrix ermöglicht.

### 4.1 Codia DMS

Codia DMS bietet auf Basis des d3.ecm von d.velop spezialisierte Lösungen im eGovernment Umfeld für öffentliche Verwaltung und Hochschulen mit den Themen Scannen Klassifizierung, Rechnungs- und Eingangspostverarbeitung, eAkte und Archivierung.

### 4.2 InPunkto

InPunkto spezialisiert auf Dokumenten Dienstleistungen im SAP Umfeld mit den Themen automatische Erfassung & Verarbeitung, Workflow, eAkte und Archivierung.

### 4.3 Übersicht

Thema	Codia	InPunkto	TaARs
Automatisierte Klassifizierung	J	J	N
Attributierung semantisch	J	kA	N
fachlich	J	J	J
Automatisierte Attributierung	N	N	J
Steuerung	lastabhängige Aufgabenverteilung über Workflowsystem	kA	Priorisierung
(Rechnungs) Workflow	d3ecm	SAP Workflow	freie Wahl
Export	d3.ecm	SAP	Rohexport als xml

## Kapitel 5

# Alleinstellungsmerkmale

Die Markt- und Domänenanalyse offenbart (hoffentlich!!!) eine Marktlücke für Firmen mit

1. mit einem bestehenden Extrahierungsprozess und optionalem Nacherfassungsprozess
2. mit einem bestehenden Buchungs-, Kontierung- oder Archivsystem
3. Workflow
  - (a) mit bestehendem Workflowsystem oder
  - (b) ohne Bedarf für ein vollintegratives Workflowsystem
4. mit Bedarf für eine Automatisierung

**Slim Clients**

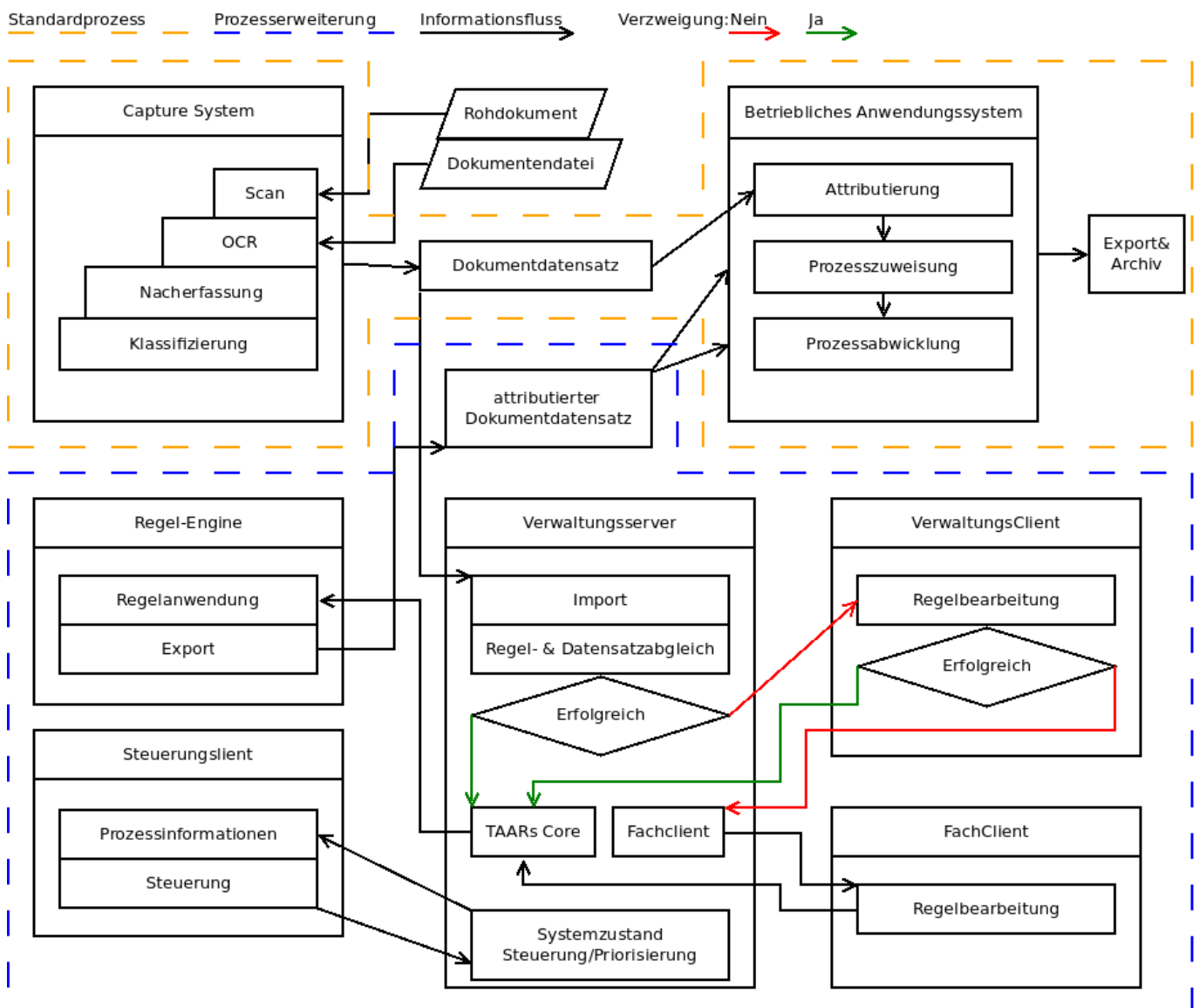
**Automatisierte Attributierung**

**freie Wahl des Workflowsystem** Für ein Workflowsystem müssen

# Kapitel 6

## Prozess

Ausgehend von ??Domaenenrecherche, Prozess eine Darstellung eines exemplarischen Verarbeitungsprozesses und die Intergration des zu entwickelnden TaARs in diesen.



# Kapitel 7

## Kommunikationsmodel

### 7.1 Akteure

#### Verwaltungsserver

- übergibt Rohdaten und Regeln an Regel-Engine
- gibt Informationen zum Systemzustand an Steuerungsclient
- erhält Priorisierungsinformation von Steuerungsclient
- übergibt priorisiertes Geschäftsobjekt an Verwaltungsclient

#### Steuerungsclient

- erhält Information vom Verwaltungsdienst zum Systemzustand
- teilt Verwaltungsdienst die Priorisierung mit

#### Verwaltungsclient

- erhält aktuell priorisierte Geschäftsobjekte vom Verwaltungsdienst
- gibt vollständiges Geschäftsobjekt an Verwaltungsdienst zurück
- gibt unvollständiges Geschäftsobjekt an Fachclient (über Verwaltungsdienst)

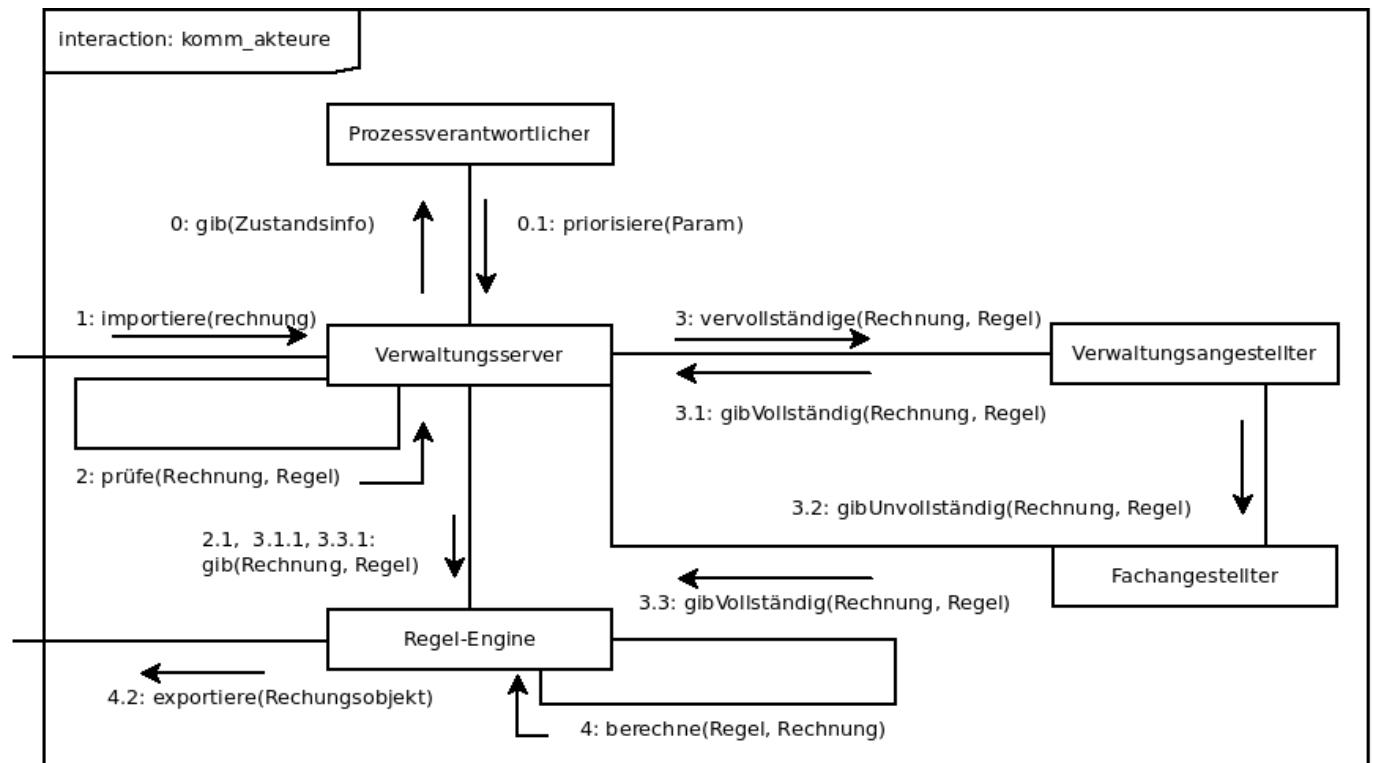
#### Fachclient

- erhält unvollständiges Geschäftsobjekt von Verwaltungsclient (über Verwaltungsdienst)
- gibt vervollständigtes Geschäftsobjekt an Verwaltungsdienst zurück

#### Regel-Engine

- erhält Geschäftsobjekt und Regel vom Verwaltungsdienst

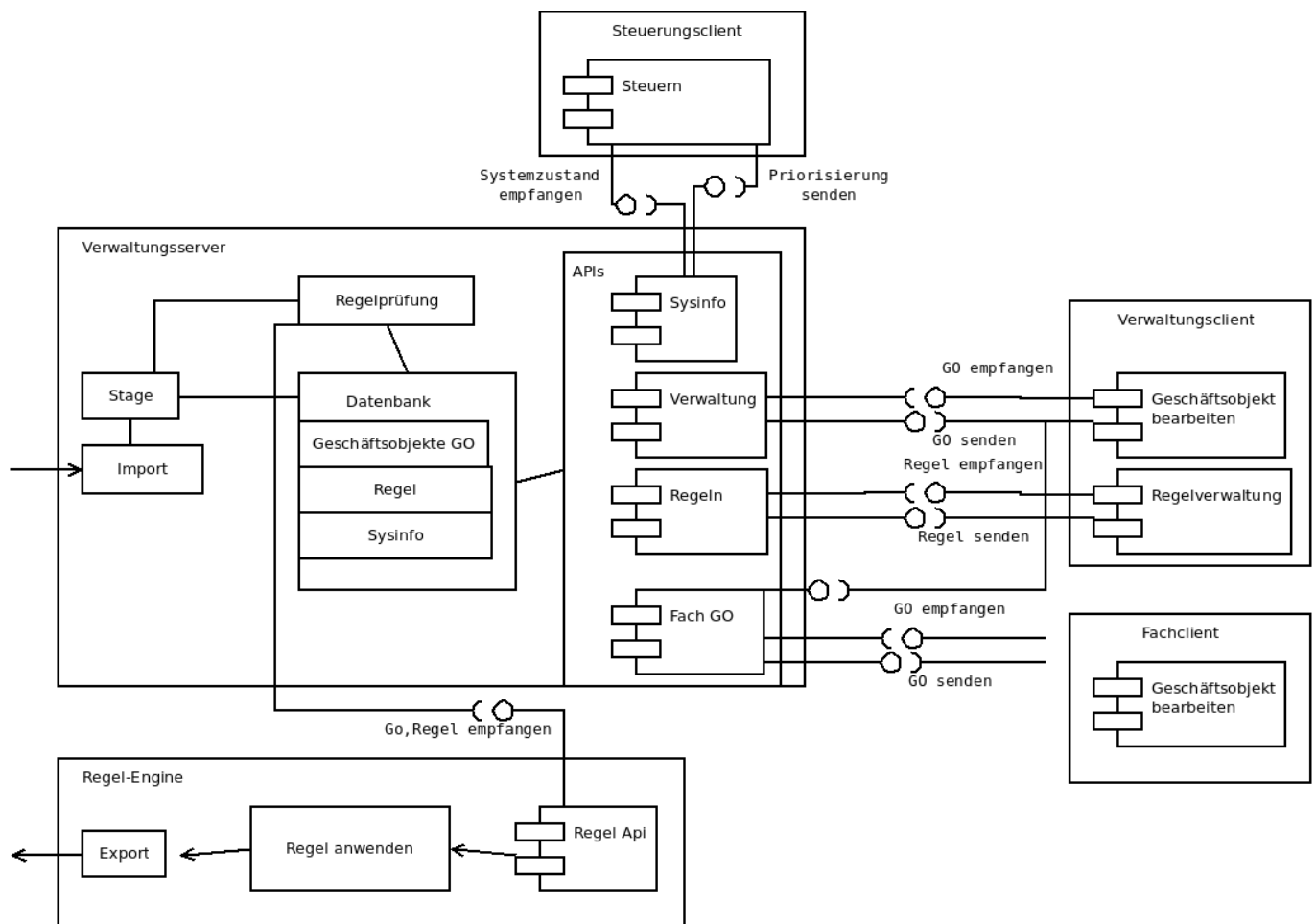
## 7.2 Kommunikationsdiagramm



# Kapitel 8

## Architekturdiagramm

### 8.1 Variante 1: API orientiert



### 8.2 Variante 2: Nachrichten orientiert



# Kapitel 9

## Risiken

### 9.1 Architektur

**Zugriffsstruktur & Steuerung** Es muss eine Struktur gefunden werden mit der die folgende Funktionalitäten unabhängig voneinander realisiert werden können:

1. Zugriff des Verwaltungsclients auf Geschäftsobjekte
2. Übergabe der Geschäftsobjekte von Verwaltungsclient an den Fachclient
3. Steuerung der Priorisierung der Geschäftsobjekte

#### **Steuerungsclient**

1. Kommunikation zu Steuerungskomponente im Verwaltungsdienst
2. Implementierung von Interaktionsparadigmen (?)

**Regel-Engine** Es muss eine Regelform entwickelt werden die die strategischen Ziele 1 + 2 erfüllt. Definition und Speicherung des Regelobjekts sowie Anwendung der Regeln auf Geschäftsobjekte.

# Kapitel 10

## Proof of Concepts

### 10.1 Definition

#### 10.1.1 Architektur

**Zugriffstruktur**

#### 10.1.2 Regel-Engine

Ein simpler Poc der folgenden Ansprüchen erfüllt

#### 10.1.3 mobiler Client

Bezug: ??Risiken 2.1 Erfolgskriterien:

- ruft Info api des Verwaltungsdienstes auf
- erhält Informationen zum Systemzustand
- stellt diese Informationen dar
- ruft Steuerungs api mit Priorisierungsparameter auf

Alternative:

Steuerungsfunktion als Browserbasierten Client umsetzen

#### 10.1.4 technisch: Steuerung Dienst

#### 10.1.5 •

### 10.2 Durchführung

## **Kapitel 11**

# **Methodischer Rahmen**

## Kapitel 12

# Projektplan

Der Projektplan wird git repo : EISWS1516<sub>Howe</sub>*projektplan.ods* geführt.

# Kapitel 13

## Projektbegründungen

### 13.1 Implementierung

Die Entscheidung der Implementierungsumgebung wird aus den strategischen Zielen 2.1 sowie Punkt 3 der Kursziele abgeleitet, der da lautet:

Für die Bewerbungen in Unternehmen oder an Hochschulen ist heute oft neben einer guten Abschlussnote auch das Vorstellen einer anspruchsvollen, gut ausgeführten Projektarbeit ein wesentliches Erfolgskriterium. Das Praktikum hat das Ziel, den Studierenden die Möglichkeit zu geben, eine solche Arbeit zu erstellen oder zumindest einen ersten signifikanten Zwischenschritt bei Erstellung einer solchen Projektarbeit zu erreichen.

Daraus folgt die Erkenntnis, dass eine fachliche und technologische Annäherung des Projekts an den antizipierten beruflichen Kontext das Ausmaß der Zielerfüllung des Kurses erhöht.

**beruflicher technologischer Kontext** Im beruflichen Kontext wird für Windows Desktop und Windows Server im Stack .NET, c#, MSSql entwickelt.

**Risiken** Eine Entwicklung im obg. Kontext würde folgende Nachteile mit sich bringen:

1. fehlende Unterstützung bei Implementierung durch Kursbetreuer
2. fehlende Portierbarkeit der Komponenten
3. ...

#### Chancen

1. höhere Bewegungssicherheit im beruflich relevanten technologischen Kontext
2. Wettbewerbsvorteil durch Erwerb technologischer Kompetenzen 'abseits der Masse'

**Entscheidung** Daraus folgt die Entscheidung, im beschriebenen technologischen Kontext zu implementieren. Es bleibt jedoch der Vorbehalt, bei Bedarf einzelne Systemkomponenten in einem anderen Kontext zu implementieren.

### 13.2 Objektbereich

Aus der ?? sowie Punkt 1 und 2 der Strategisch folgt die Entscheidung, dass im Rahmen dieses Projekts der Objektbereich auf den Dokumenttyp Rechnung eingegrenzt wird.

### 13.3 Nutzermodelle

Auf die Rückmeldung der Betreuer, siehe vom 12.10.2015, Punkt Rückmeldung, den MCI relevanten Anteil zu erhöhen, wird folgendermaßen reagiert:

**Variante 1: Nutzermodelle spezialisieren** Eine Spezialisierung der Nutzermodelle würde einen Konflikt mit den strategischen Zielen bedeuten 2.1 die eine möglichst breite wirtschaftliche Anwendungsdomäne anvisieren. Zudem werden auch in spezialisierteren Anwendungsgebieten allgemeingültige Objektbereiche verarbeitet, siehe ??quelle: codia an hochschulen.

**Variante 2: Nutzungsmodelle erhöhen** Um den MCI Anteil über zusätzliche Funktionen(?)/Nutzungsmodelle zu erhöhen muss ein Modell ausserhalb derer in Anwendungsdomaene Nutzung betrachteten gefunden werden. Dies böte die Chance eine größere Bandbreite der Nutzerinteraktion und Nutzungskontexte zu bearbeiten, bürge jedoch die Risiken des zusätzlichen Entwicklungsaufwandes für einen andersartigen Client sowie ggf. erhöhte Planungsunsicherheit aufgrund der Nutzung von weniger geübten Technologien.

**Entscheidung** Die Entscheidung fällt zugunsten der Variante 2, die in Systembeschreibung umschrieben und in die Artefakte aufgenommen wird.

## 13.4 Vorgehensmodell

# Kapitel 14

## Quellen

**Domänenrecherche** - <https://de.wikipedia.org/wiki/Dokumentenklasse>

- <http://www.officemanager.de/glossar/dokumentenmanagement.html>

- [http://dbis.eprints.uni-ulm.de/782/1/Diplomarbeit\\_Final.pdf](http://dbis.eprints.uni-ulm.de/782/1/Diplomarbeit_Final.pdf)

*http://www.project – consult.de/ecm/content/capture,*

**Marktcherche** - isoft: [http://www.i-soft-systemhaus.de/fileadmin/templates/pdf/isoft\\_workflowRE.pdf](http://www.i-soft-systemhaus.de/fileadmin/templates/pdf/isoft_workflowRE.pdf)

– codia : [http://www.codia.de/codia/files/dms\\_in\\_hochschulenv1.pdf](http://www.codia.de/codia/files/dms_in_hochschulenv1.pdf)