



# Literatur-Review

**Data Governance Aufgaben Herausforderungen und Lösungen  
zur Sicherung der Datenqualität.**

über das Thema

**Vorgehensweise zur Bestimmung der Datenqualität  
(Datenqualitätsprozess und DQ-Assessment)**

**Autor:** Vedad Hamamdžić  
vhamamdžić@stud.hs-heilbronn.de

**Prüfer:** Dipl.-Inf. Thomas Schäffer

**Abgabedatum:** 09.01.2015

## **I Kurzfassung**

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

## **Abstract**

Das ganze auf Englisch.

## II Inhaltsverzeichnis

<b>I</b>	<b>Kurzfassung</b>	<b>I</b>
<b>II</b>	<b>Inhaltsverzeichnis</b>	<b>II</b>
<b>III</b>	<b>Abbildungsverzeichnis</b>	<b>III</b>
<b>IV</b>	<b>Tabellenverzeichnis</b>	<b>IV</b>
<b>V</b>	<b>Listing-Verzeichnis</b>	<b>IV</b>
<b>VI</b>	<b>Abkürzungsverzeichnis</b>	<b>V</b>
<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Begriffserklärung</b>	<b>2</b>
<b>3</b>	<b>Methodik</b>	<b>3</b>
3.1	Suchprotokoll . . . . .	5
<b>4</b>	<b>Faktor Datenqualität</b>	<b>5</b>
<b>5</b>	<b>Technische Faktoren</b>	<b>5</b>
<b>6</b>	<b>Faktor Organisation</b>	<b>5</b>
6.1	Bilder . . . . .	5
6.2	Tabellen . . . . .	6
6.3	Auflistung . . . . .	7
6.4	Listings . . . . .	7
6.5	Tipps . . . . .	7
<b>7</b>	<b>Kapitel</b>	<b>8</b>
7.1	Unterkapitel . . . . .	8
7.2	Unterkapitel . . . . .	8
<b>8</b>	<b>Kapitel</b>	<b>9</b>
8.1	Unterkapitel . . . . .	9
8.2	Unterkapitel . . . . .	9
<b>9</b>	<b>Kapitel</b>	<b>10</b>
9.1	Unterkapitel . . . . .	10
9.2	Unterkapitel . . . . .	10
<b>10</b>	<b>Quellenverzeichnis</b>	<b>11</b>
	<b>Anhang</b>	<b>I</b>
<b>A</b>	<b>GUI</b>	<b>I</b>

### III Abbildungsverzeichnis

Abb. 1	Fünf Phasen der Reviewforschung . . . . .	3
Abb. 2	OSGi Architektur . . . . .	6

## **IV Tabellenverzeichnis**

Tab. 1	Beispieltabelle . . . . .	5
Tab. 2	Beispieltabelle . . . . .	6

## **V Listing-Verzeichnis**

Lst. 1	Arduino Beispielprogramm . . . . .	7
--------	------------------------------------	---

## **VI Abkürzungsverzeichnis**

**OSGi** Open Service Gateway initiative

# 1 Einleitung

Daten sind im wirtschaftlichen Kontext betrachtet, eine wichtige Resouce. Doch ungeachtet des hohen Stellenwerts sind Daten von niedriger Qualität in Unternehmen allgegenwärtig. Gerade diese, mangelhaften Daten können wirtschaftliche Konsequenzen mit sich bringen. Nach dem garbage-in-garbage-out Prinzip zieht sich die Problematik durch sämtliche Prozesse, die diese Daten nutzen. Doch was sind Daten eigentlich? Nach der Definition von Rehäuser und Krcmar :

Informationen sind Daten, die in einen Kontext – also einen Problemzusammenhang – gestellt wurden (vgl. Rehäuser u. Krcmar 1996, S. 4, [RK96]).

Somit werden Buchstaben, Ziffern und Sonderzeichen durch Ordnungsregeln (einen Code oder eine Syntax) zu Daten. [Nor11] Doch Datenqualität beginnt nicht erst bei dem Erfassen der Daten. Auch andere Faktoren spielen eine Rolle wie z.B. Modellierung der Datenbanken aber auch Organisatorische Aspekte sind nicht zu vernachlässigen. Um Datenqualität im Unternehmensumfeld zu gewährleisten ist es sicherlich nicht falsch das Unternehmen als individuelles Ganzes zu betrachten. Denn jedes Unternehmen hat aufgrund der einzigartigen Bedürfnisse ihres Systems mit verschiedenen Problemen zu kämpfen. Die Heterogenität der Standards in der IT Branche erfordert diese Betrachtungsweise. Jedoch soll diese Review eine Data Governance definieren die allgemein gültige Richtlinien aufzeigt. Es stellt sich die Frage, Welche Anforderungen an ein Data Governance sollte ein Unternehmen hinsichtlich einer hohen Datenqualität erfüllen? Um Antworten auf die Frage zu finden werde ich Faktoren extrahieren die Einfluss auf Qualität von Daten haben. Wie bereits erwähnt sind da Organisatorische und Technische Aspekte zu untersuchen. Anfangs ist es jedoch essenziell klar zu stellen, was unter hoher Datenqualität zu verstehen ist. Daher wird die Review auch so aufgebaut sein das zunächst untersucht wird was die Literatur über Qualitätsmerkmale von Daten Definiert. Im weiteren Verlauf werden Publikationen untersucht welche die Vorgehensweisen zur Steigerung der Qualität im Technischen und Architektonischen Kontext beschreibt. Als letzter Faktor der Einfluss auf die Qualität von Daten hat, werden Kriterien an das Informationsmanagement und Organisation untersucht. Durch das Rubrizieren ist man in der Lage die Auswahl für die Literatur besser einzugrenzen. Es trennt aber auch klar in welchen Fachbereichen gewisse Regeln gelten sollen und Definiert somit auch Richtlinien für die einzelnen Bereiche eines Unternehmens.

## 2 Begriffserklärung

Im Folgenden werden die Grundbegriffe Information, Daten, Datenqualität, Data Governance eingeführt und definiert.

### Information und Daten

Oft werden die Begriffe Daten, Information und Wissen als Synonyme verwendet, doch dem ist aber nicht so. Daten entstehen erst wenn man Zeichen in einen definierten, strukturierten Zusammenhang gebracht hat, erst dann kann man von **Daten** sprechen. Ein Beispiel "525" sind drei Zahlen versehen ich sie noch mit einem Komma 52,50 \$ und einem Dollar Zeichen so wird es zur **Information**. Gibt man noch einen Kontext dazu wie z.B. Rohölpreis in Barel hat man Wissen.

### Datenqualität

Daten von schlechter Qualität enthalten Datenfehler, Redundanzen, fehlende Werte, falsche Formatierungen usw. Shawn Turner definierte im Jahre 2002 den Begriff Datenqualität

Die Datenqualität ist die Fitness der Daten für alle Zwecke, die sie benötigen. Messdatenqualität erfordert ein Verständnis aller beabsichtigten Verwendungszwecke für diese Daten. [T<sup>+</sup>02]

Damit meint Turner die Eignung der Daten für die weitere Nutzung, ganz unabhängig davon ob diese durch Anwendungen geschieht oder Personen.

### Data Governance

Für Data Governance gibt keine einheitliche Definition, eher eine Abgrenzung des Begriffs.

Data Governance ist der Rahmen für Datenqualitätsmanagement und legt fest, welche Rollen mit welchen Zuständigkeiten die Aufgaben des Datenqualitätsmanagement übernehmen. Häufig konkretisiert sich Data Governance in Richtlinien und Vorgaben für den Umgang und die Pflege von Daten gemäß den betriebswirtschaftlichen Zielsetzungen des Unternehmens.[Hil08]



### 3 Methodik

Die Autoren Webster und Watson (2002) haben eine Definition der Literaturanalyse veröffentlicht. Sie definierten diese Forschungsmethode als:

„solide Grundlage mit welcher es möglich sei, die Menge der existierenden Erkenntnisse darzulegen und im Anschluss daran aufzuzeigen in welchen Bereichen weitere Forschung angestrebt werden sollte“ [WW02]

Die Literaturanalyse als Forschungsmethode lässt sich nach dem Handbuch der Reviewforschung [HED93] methodisch in fünf Phasen unterteilen (Abbildung 1). Auch Peter Fettke erwähnt diese Phasen in seinem Artikel State-of-the-Art des State-of-the-Art <sup>1</sup>. Im Weiterem Verlauf der Analyse werde ich die Phasen der Reviewforschung untersuchen und hinsichtlich meiner Forschungsfrage abarbeiten.

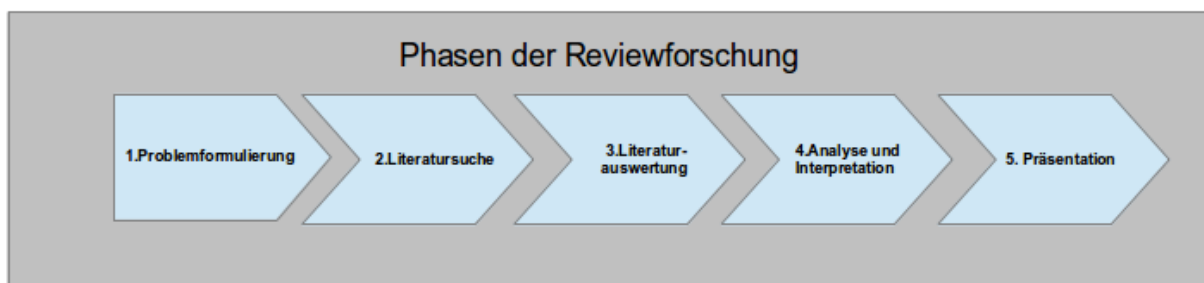


Abbildung 1:

Fünf Phasen der Reviewforschung

Während der ersten Phase wird die Frage, die durch das Review zu beantworten ist präzise ausformuliert. Das auswerten der Literatur ist in der zweiten Phase beschrieben. Darin geht es Primär darum für die Fragestellung geeignete Literatur zu recherchieren. Darauf folgt die Identifizierung von Relevanter Literatur welche man dann in der vierten Phase Analysiert und Interpretiert. Das Präsentieren der Ergebnisse ist die letzte der fünf Phasen. Für diese Analyse werden Primär Quellen in spezialisierte Suchmaschinen für wissenschaftlich Publikationen gesucht. Während der Untersuchung und Ausarbeitung dieser Analyse wurden ausschließlich folgende Suchmaschinen genutzt.

- [link.springer.com](http://link.springer.com)
- [scholar.google.de](http://scholar.google.de)
- [boss.bsz-bw.de/HSHN/](http://boss.bsz-bw.de/HSHN/)

zur Auswahl der in dieser Arbeit benutzten wissenschaftlichen Zeitschriften gehören die vom Springer-Verlag herausgegebene Zeitschrift Informatik Spektrum sowie die Zeitschrift

---

<sup>1</sup>Vgl. [Fet06]

Wirtschaftsinformatik sowie "Journal of management information systems". Diese Zeitschriften sind auf der WI-Journalliste 2008 zu finden. Diese Liste soll als Orientierung dienen. Ausgearbeitet von zwei Arbeitskreisen deren Mitglieder universitäts Professoren aus Österreich, Deutschland und der Schweiz sind [WiL08]. Um zu verdeutlichen welche Herausforderungen in der Praxis existieren werden auch Artikel aus der Fachzeitschrift Controlling Management und die Untersuchung miteinbezogen. Der zeitlichen Intervall zur Auswahl der Literatur wird auf 1990 bis 2014. Weil seit Mitte der 1990er Jahre Systeme, in einem Unternehmen zur Entscheidungsfindung herangezogen werden.<sup>2</sup> Für der Art von Entscheidungen ist eine hohe Datenqualität von enormer Wichtigkeit. Da lässt sich in dieser Zeitspanne kurz vor Beginn der Nutzung Entscheidungsunterstützender Systeme bis heute Relevante Literatur finden.

Die Keywords welche zur Recherche in den oben genannten Suchmaschinen genutzt werden sind dem Kontext des Fachbereichs einzuordnen.

Keywords Rubrik Datenqualität

- –Datenqualität - data quality
- –Data Governance

Keywords Rubrik Technik & Architektur

- –Data Warehouse
- –Datenbankmanagementsystem
- –Daten bereinigung - Data cleaning

Keywords Rubrik Informationsmanagement und Organisation

- Informationsmanagement
- Datenqualitätsmanagement

Die Suchbegriffe sind den Rubriken zuzuordnen, für die Datenqualität sowie Data Governance unter der Rubrik Datenqualität gilt es zu untersuchen welche Definitionen und Ansätze es bereits gibt. Der Begriff Data Warehouse soll aufzeigen ob Unterschiede in der Modellierung sowie der Architektur eines Systems Einfluss auf die Datenqualität haben. Datenbankmanagementsystem als Suchbegriff soll Klarheit schaffen ob die Wahl des DMS eine Rolle hinsichtlich der Datenqualität spielt. Aber auch das Bereinigen von Daten soll untersucht werden um Nachhaltigkeit zu gewährleisten. Die Rubrik Informationsmanagement und Organisation soll den Faktor Mensch repräsentieren, und aufzeigen in wie weit Datenqualität abhängig vom Menschen ist.

---

<sup>2</sup>[HW05]

### 3.1 Suchprotokoll

Der Suchbegriff Datenqualität ergab 10 Ergebnisse wobei die suche durch Angabe des Zeitintervalls von 1990 bis 2014 und nur im Titel gefiltert wurde. Diese sind soweit als Grundeinstellung für jede weitere suche bei Springerlink zu sehen. Datenqualitätsmanagement ergab 16 Treffer, 17 Artikel wurden bei dem Suchbegriff Datenbankmanagementsystem angezeigt. Weiter suchen ergaben folgendes:

Suchbegriff	Anzahl	Suchmaschine
Data Governance	1	Springerlink
Data Warehouse	7	Springerlink
Datenbereinigung	32	Springerlink
Informationsmanagement	15	Springerlink

Tabelle 1: Beispieltabelle

Es werden nur Publikationen berücksichtigt welche Beschreiben wie Qualität der Daten verbessert werden

## 4 Faktor Datenqualität

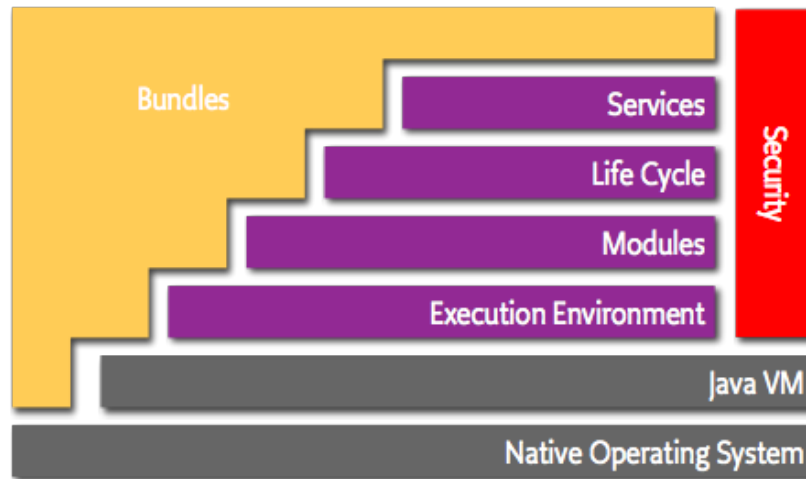
Die Qualität von Daten wird von verschiedensten Faktoren bestimmt, laut Redman die Fähigkeit, die Anforderungen der beabsichtigten Nutzung in einer bestimmten Situation zu erfüllen („fitness for use“) [Red96] in Kombination mit den Qualitätsmerkmalen nach Wang Strong, welche die Merkmale Klassifizieren und somit Redmans Definition stützen.

## 5 Technische Faktoren

## 6 Faktor Organisation

### 6.1 Bilder

Zum Einfügen eines Bildes, siehe Abbildung 2, wird die *minipage*-Umgebung genutzt, da die Bilder so gut positioniert werden können.

Abbildung 2: OSGi Architektur<sup>3</sup>

## 6.2 Tabellen

In diesem Abschnitt wird eine Tabelle (siehe Tabelle 2) dargestellt.

Name	Name	Name
1	2	3
4	5	6
7	8	9

Tabelle 2: Beispieltabelle

---

<sup>3</sup>Quelle: <http://www.osgi.org/Technology/WhatIsOSGi>

### 6.3 Auflistung

Für Auflistungen wird die *compactitem*-Umgebung genutzt, wodurch der Zeilenabstand zwischen den Punkten verringert wird.

- Nur
- ein
- Beispiel.

### 6.4 Listings

Zuletzt ein Beispiel für ein Listing, in dem Quellcode eingebunden werden kann, siehe Listing 1.

---

```
1  int ledPin = 13;
2  void setup() {
3      pinMode(ledPin , OUTPUT);
4  }
5  void loop() {
6      digitalWrite(ledPin , HIGH);
7      delay(500);
8      digitalWrite(ledPin , LOW);
9      delay(500);
10 }
```

---

Listing 1: Arduino Beispielprogramm

### 6.5 Tipps

Die Quellen befinden sich in der Datei *bibo.bib*. Ein Buch- und eine Online-Quelle sind beispielhaft eingefügt. [Vgl. [buc], [onl]]

Abkürzungen lassen sich natürlich auch nutzen (Open Service Gateway initiative (OSGi)). Weiter oben im Latex-Code findet sich das Verzeichnis.

## 7 Kapitel

Lorem ipsum dolor sit amet.

### 7.1 Unterkapitel

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

### 7.2 Unterkapitel

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

## 8 Kapitel

Lorem ipsum dolor sit amet.

### 8.1 Unterkapitel

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

### 8.2 Unterkapitel

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

## 9 Kapitel

Lorem ipsum dolor sit amet.

### 9.1 Unterkapitel

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

### 9.2 Unterkapitel

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.



## 10 Quellenverzeichnis

[buc]

- [Fet06] FETTKE, Dipl-Wirt-Inf P.: State-of-the-Art des State-of-the-Art. In: *Wirtschaftsinformatik* 48 (2006), Nr. 4, S. 257–266
- [HED93] HEDGES, HARRIS COOPER LARRY V.: RESEARCH SYNTHESIS AS A SCIENTIFIC ENTERPRISE. In: *The Handbook of Research Synthesis* (1993), S. 3
- [Hil08] HILDEBRAND, Knut: *Daten-und Informationsqualität: auf dem Weg zur Information excellence*. Springer-Verlag, 2008
- [HW05] HUMM, Bernhard ; WIETEK, Frank: Architektur von Data Warehouses und Business Intelligence Systemen. In: *Informatik-Spektrum* 28 (2005), Nr. 1, 3-14. <http://dx.doi.org/10.1007/s00287-004-0450-5>. – DOI 10.1007/s00287-004-0450-5. – ISSN 0170-6012
- [Nor11] NORTH, Klaus: *Wissen in Organisationen*. Springer, 2011. – S. 35–68.

[onl]

- [Red96] REDMAN, Thomas: *Data quality for the information age*. Boston : Artech House, 1996. – ISBN 0890068836
- [RK96] REHÄUSER, Jakob ; KRCMAR, Helmut: *Wissensmanagement in Unternehmen*. Lehrstuhl für Wirtschaftsinformatik, Univ. Hohenheim, 1996
- [T<sup>+</sup>02] TURNER, Shawn u. a.: Defining and measuring traffic data quality. In: *Proceedings of the Traffic Data Quality Workshop, Washington, DC*, 2002
- [WiL08] WI-Orientierungslisten. In: *WIRTSCHAFTSINFORMATIK* 50 (2008), Nr. 2, 155-163. <http://dx.doi.org/10.1365/s11576-008-0040-2>. – DOI 10.1365/s11576-008-0040-2. – ISSN 0937-6429
- [WW02] WEBSTER, Jane ; WATSON, Richard T.: Analyzing the past to prepare for the future: Writing a literature review. In: *Management Information Systems Quarterly* 26 (2002), Nr. 2, S. 3

## **Anhang**

### **A GUI**

Ein toller Anhang.

#### **Screenshot**

Unterkategorie, die nicht im Inhaltsverzeichnis auftaucht.

# Erklärung

Hiermit versichere ich, dass ich meine Abschlussarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Datum:

.....

(Unterschrift)