

## **Exposé**

Bereitstellung offener Daten aus öffentlichen Einrichtungen  
in Deutschland durch Content-Management-Systeme  
am Beispiel von Hochschuldaten und TYPO3

## **Bachelor-Arbeit**

zur Erlangung des Grades Bachelor of Science  
im Online-Studiengang Medieninformatik  
der VFH an der Technischen Hochschule Brandenburg

### **vorgelegt von**

Marco Habich  
Nienburg / Weser

Betreuer: Sebastian Kreideweiß, M.Sc.

Zweitgutachter: Prof. Dr. Vera Meister

Nienburg / Weser, den 3. April 2023

## Inhaltsverzeichnis

Medienverzeichnis .....	III
Abkürzungsverzeichnis .....	III
Glossar .....	III
1 Problemstellung .....	1
2 Erkenntnisinteresse .....	1
3 Fragestellung .....	2
4 Ziele und Hypothesen .....	3
5 Methode .....	3
6 Material .....	4
7 Gliederungsentwurf .....	5
8 Zeitplan .....	5
Literaturverzeichnis .....	6

## Anhang

Im Exposé sind keine Anhänge vorhanden.

## Medienverzeichnis

[Med\_T3] TYPO3, Get TYPO3 11.5.25  
Download am 25.03.2023  
<https://get.typo3.org/11.5.25/zip>

## Abkürzungsverzeichnis

CMS	Content-Management-System
CSS	Cascading Style Sheets
CSV	Comma-separated Values
HTML	Hypertext Markup Language
JS	JavaScript
JSON	JavaScript Object Notation
KI	Künstliche Intelligenz
PDF	Portable Document Format
PHP	Hypertext Preprocessor
RDF	Resource Description Framework
THB	Technische Hochschule Brandenburg
UC	Use Case
WWW	World Wide Web
XML	Extensible Markup Language

## Glossar

Extension	Komponente, um TYPO3 zu erweitern
localhost	Domainname von einem lokalen Rechner
macOS	Betriebssystem
Redakteur	Rolle/Nutzer in TYPO3, die/der Seiteninhalte bearbeiten kann
TYPO3	CMS, frei verfügbar
Use Case	deutsch: Anwendungsfall, beschreibt Aktionen, bzw. mögliche Szenarien, um ein fachliches Ziel zu erreichen

# 1 Problemstellung

Gegeben ist der Umstand, dass eine Vielzahl von CMS am Markt existieren (1.500 laut cmsmatrix.org). Die Entstehungsgeschichte einiger Systeme reicht auf das Ende der 1990er und Anfang der 2000er Jahre zurück.

Heutige Anforderungen an ein CMS verlangen die Datenhaltung getrennt vom Ausgabekanal und deren Aufbereitung in verschiedensten offenen Formaten. Nicht zuletzt durch Gesetze [BMI\_Od] formuliert, sollen Rohdaten bereitgestellt werden und die Weiterverarbeitung durch dritte Anwendungen leichter ermöglichen.

Schwerpunkt der Arbeit soll auf Daten aus öffentlichen Einrichtungen, z.B. Hochschulen, in Deutschland liegen. TYPO3 soll dabei, aufgrund der Verbreitung [CMS\_cen], als CMS beispielhaft herangezogen werden.

Die These ist, dass in solchen Systemen ein erheblicher „Datenschatz“ vorliegt, der jedoch **nicht** in eindeutiger und maschinenlesbarer Form für dritte Anwendungen weiterverarbeitbar ist. Daten können somit nicht in einem Knowledge Graph für (KI-) Anwendungen bereitgestellt werden.

Anhand eines vorliegenden Konzeptes [ResGate\_Me] könnte der UC „Hochschuldaten“ bearbeitet werden. Als weitere Referenz dient eine Master-Thesis [ResGate\_Kr] aus 2017.

## 2 Erkenntnisinteresse

Das Bestreben ist, einen optimierten Weg zu finden, der die Bereitstellung von offenen Daten erleichtert. Durch den Abruf dieser Daten durch Dritte, z.B. in leicht zugänglichen Anwendungen, profitiert die Gesellschaft.

### 3 Fragestellung

- a) Welcher Datenschatz liegt vor? Was macht die jetzige Weiterverarbeitung so schwierig. These: Vieles liegt in Tabellenform in PDF, Excel oder HTML ohne einheitlich verarbeitbare Struktur vor.
- b) Was sind maschinenlesbare Datenformate (z.B. JSON, XML, RDF, n3, turtle, SPARQL-Endpoints statt HTML, CSV, Excel, PDF)?
- c) Welche Bedeutung hat digitale Souveränität (z.B. PDF = Adobe, Excel = Microsoft) dabei?
- d) Wie kann ein CMS wie TYPO3 den Herstellungsprozess dieser Daten beeinflussen? These: Ein „Digital first“-Ansatz muss als grundsätzliches Mindshift Einzug halten. Bisläng wird oft die gedruckte Broschüre eins zu eins als PDF digitalisiert, in ein CMS hochgeladen und wichtige Daten im besten Fall per Hand individuell aufbereitet. Unmöglich, hierauf nachhaltige Anwendungen zu entwickeln.
- e) Was und wie machen es andere CMS (z.B. WordPress, Drupal)?
- f) Was muss eine Komponente bzw. Lösung leisten? Architekturfragen:
  - a. Kann das in einem CMS wie TYPO3 gelöst werden?
  - b. Braucht man eine zusätzliche Software wie ckan.org
- g) Welche Standardisierung der Konzepte (schema.org, etc.) gibt es oder müssen geschaffen werden?
- h) Was wäre ein UC, um die o.a. Fragen anhand eines Beispiels durchzuspielen? Beispiele:
  - I. in einer Pandemie: Wie können Gesundheitsämter Infektionszahlen bereitstellen, sodass Datenjournalisten, App-Hersteller, etc. darauf zugreifen können?
  - II. an Hochschulen (Fortführung): Wie können strukturierte Hochschuldaten, wie Studienangebote [ResGate\_Me], für eine zentrale Anwendung dargestellt werden?
- i) Insgesamt: Wie können Daten in CMS angereichert und an Datenportale wie [daten.berlin.de/](https://daten.berlin.de/), <https://www.datenadler.de/>, <https://www.govdata.de/> oder <https://data.europa.eu> als Drittsysteme aus einem TYPO3 System weitergereicht werden?

## 4 Ziele und Hypothesen

Im Rahmen der Bachelor-Arbeit soll identifiziert werden, wie und womit die Bereitstellung offener Daten optimiert werden kann. Dabei liegt das Hauptaugenmerk auf Hochschuldaten, die mittels TYPO3 verwaltet werden und der Öffentlichkeit im WWW zur Verfügung stehen.

Der Redakteur einer TYPO3-Instanz soll bei der Nutzung so wenig wie möglich Mehraufwand haben, um die offenen Daten bereitzustellen. Weitere Vorteile, die durch einen etwaigen Prozess geschaffen werden, könnten sein:

- Automatismus
- Datentransfer / Schnittstellen
- Dritte profitieren von diesen Daten

## 5 Methode

Es wird Zeit in eine Recherche und deren Auswertung investiert, um einen Wissensstand, bezugnehmend auf das Thema, aufzubauen. Diese Maßnahmen werden, anhand der Ergebnisse, gegebenenfalls iterativ erfolgen. Im Anschluss werden theoretische Möglichkeiten, unter anderem mittels Pro und Contra, aufgezeigt. Sollte eine dieser Theorien praxisnah umzusetzen sein und den Rahmen der Bachelor-Arbeit nicht überschreiten, wird ein Prototyp angestrebt.

Bei den Recherchen wäre es hilfreich auf eine bereits vollständig implementierte und seit Jahren genutzte TYPO3-Instanz zuzugreifen. Dabei soll der Zugriff lediglich als „lesender Redakteur“ erfolgen, um zu erfahren, wie der momentane Umgang in existenten Systemen ist.

## 6 Material

Wenn ein Praxisbezug besteht, der auf dem Gerät des Autors erledigt werden muss, z.B. für den möglichen Bau einer Extension, ist zur Erfüllung der Aufgaben TYPO3 [Med\_T3] vorgesehen.

Die Umgebung, auf macOS im „localhost“-Modus, ist in den Folgepunkten aufgeführt. Je nach Ansatz zieht der Autor weitere Werkzeuge und Sprachen in Betracht, u.a. HTML, CSS, JS, Frameworks, Python.

### TYPO3

- Version 11.5.25
- ggf. Version 12.2.x

### Visual Studio Code

- Version 1.76.2 (Universal)
- Datum 2023-03-14T17:54:09.061Z

### System

- MacBook Pro (Retina 13“, 2012)
- 2,5 GHz Dual-Core Intel Core i5
- 8 GB 1600 MHz DDR3
- macOS Catalina (Version 10.15.7)

### MAMP

- Version 6.8 (1258)
- Apache
- PHP 8.2
- MySQL Version 5.7.39

### Browser

- Safari (Version 15.6.1)
- Firefox (Version 111.0.1 / 64-Bit)
- Chrome (Version 111.0.5563.110)
- Opera (Version 74.0.3911.218)

Für das Verfassen der Bachelor-Arbeit wird *Pages* (Version 11.0) verwendet, wobei die finale Fassung digital als PDF abgegeben wird (Betreuer, Zweitgutachter, Prüfungsamt). Etwaige Präsentationsunterlagen für das Kolloquium werden voraussichtlich mit *OpenOffice* (Version 4.1.1) gestaltet. Sollten unter den bereits genannten Tools die integrierten Hilfsmittel nicht ausreichen, wird für eine grafische Unterstützung auf *Figma Desktop App* (Version 116.7.103) oder *GIMP* (Version 2.10.34) zurückgegriffen.

## 7 Gliederungsentwurf

Die Gliederung der Bachelor-Arbeit erfolgt nach einem bewährten Muster.

<b>Vorspann</b>	<b>Kern</b>	<b>Nachspann</b>
- Inhaltsverzeichnis	- Einleitung	- Literaturverzeichnis
- Abbildungsverzeichnis	- Hauptteil	- Anhang
- Tabellenverzeichnis	- Schlussteil	- Ehrenwörtliche Erklärung
- Medienverzeichnis		
- Abkürzungsverzeichnis		
- Glossar		

## 8 Zeitplan

Folgende Angaben beinhalten teilweise iterative Aktionen (*kursiv*) und sind grob definiert, außer dem Beginn und Ende der Bearbeitungszeit, die nach bisherigem Exposé-Stand fix sind.

28.03.2023	Entwurf Exposé
31.03.2023	Besprechung Exposé
06.04.2023	Fertigstellung Exposé
11.04.2023	Anmeldung der Bachelor-Arbeit beim Prüfungsamt
15.04.2023	Beginn der Bearbeitungszeit
17.04.2023	<i>Recherche-Sammlung</i>
24.04.2023	<i>Recherche-Auswertung</i>
01.05.2023	<i>Entwurf Bachelor-Arbeit</i>
08.05.2023	<i>Praxisbezug, z.B. Extension und/oder Schnittstellen</i>
30.06.2023	<i>Korrekturlesen, auch durch Dritte</i>
14.07.2023	Abgabe Digitalversion an Betreuer, Zweitgutachter, Prüfungsamt
14.07.2023	Anfertigung von Präsentationsunterlagen für Kolloquium
15.07.2023	Ende der Bearbeitungszeit
07-08/2023	Gutachtenerstellung Betreuer, Zweitgutachter
23.08.2023	Kolloquium/Verteidigung, vorzugsweise online



## Literaturverzeichnis

- [BMI\_Od] Bundesministerium des Innern und für Heimat, Open Data.  
Abgerufen am 25.03.2023  
<https://www.bmi.bund.de/DE/themen/moderne-verwaltung/open-government/open-data/open-data-node.html>
- [CMS\_cen] Judith Kölbel, Statistische Verteilung.  
Abgerufen am 25.03.2023  
<https://cmscensus.eu>
- [ResGate\_Kr] Sebastian Kreideweiß, M.Sc., TH Brandenburg.  
Download am 25.03.2023  
<https://www.researchgate.net/project/Konzeption-und-Implementierung-einer-TYPO3-Extension-zur-semantischen-Annotation-von-Studiengaengen-auf-Web-Seiten-deutscher-Hochschulen>
- [ResGate\_Me] Prof. Dr. Vera Meister, TH Brandenburg.  
Download am 25.03.2023  
[https://www.researchgate.net/publication/326732018\\_Konzept\\_und\\_Prototyp\\_einer\\_dezentralen\\_Wissensinfrastruktur\\_zu\\_Hochschuldaten\\_fur\\_Mensch\\_und\\_Maschine](https://www.researchgate.net/publication/326732018_Konzept_und_Prototyp_einer_dezentralen_Wissensinfrastruktur_zu_Hochschuldaten_fur_Mensch_und_Maschine)