Friedrich-Schiller-Universität Jena

Lehrstuhl für Geoinformatik

WiSe 2017

Abschlussbericht GEO413: Geodatenbanken

**Erstellung einer räumlichen Datenbank zur Verwaltung von Messdaten**

Vorgelegt von:

|  |
| --- |
| Krüger, Eric |
| Studiengang: M. Sc. Geoinformatik  Matrikel: 142645 |
| Adresse: Jenergasse 15  07743 Jena |
| Email: eric.krueger@uni-jena.de  Abgabedatum: 30.11. 2017  Modulverantwortlicher: Carsten Busch |

**Inhalt**

Abbildungsverzeichnis

Tabellenverzeichnis

Abkürzungsverzeichnis

1) Aufgabenstellung

2) Methodisches Vorgehen

3) Entity Relationship Modell

4) Überführung in PostGIS Messdatenbank

5) Datenquellen und Vorbereitung

6) Datenimport

7) Überführung in die Messdatenbank

8) Notwendige Sichten

9) Zusammenfassung

10) Literatur

**Abkürzungen**

DB

DWD

ERM

TLUG

1. **Aufgabenstellung**

Ziel des Geodatenbankkurses ist die Erstellung einer PostgreSQL/PostGIS Datenbank zur Verwaltung von Messdaten und deren zugehörigen Metadaten.

Folgenden Informationen sollen in der Datenbank verwaltet werden:

Messdaten von Pegelstationen, welche Wasserstand und Durchfluss umfassen, sowie Messdaten von Klimastationen, welche Temperatur, Niederschlag, Sonnenscheindauer und max. Windgeschwindigkeit beinhalten. Dazu gehörenden sollen die Metadaten von Klima- und Pegelstationen abgespeichert werden, welche an den Stationen erhoben werden, sowie die Stammdaten dieser. Für die Qualitätssicherung sind Informationen wie der Ansprechpartner, die verantwortliche Organisation, Aussagen zu Datenqualität und den Aufnahmeverfahren vorzusehen.

In der Datenbank sind weiterhin die Daten zu den Bundesländern und deren Geometrien zu verwalten. Des Weiteren soll ein entsprechendes E/R Modell mit geeigneten Attributen entwickelt werden. Dieses Modell soll in eine PostgreSQL/PostGIS Datenbank überführt werden. Dafür können Sie SQL Skripte, die PG-Admin3 Oberfläche für die Datenmodellierung/Tabellenerstellung nutzen. Danach soll die Datenbank mit Beispieldaten der TLUG (Stationen und fünf Zeitreihen zu Wasserstand/Durchfluss) und des DWD (Zeitreihen zu Klimadaten Temperatur, Niederschlag, Sonnenscheindauer, max. Windgeschwindigkeit) befüllt werden. Überlegen Sie sich bei dem Design, inwieweit sich die Möglichkeit der Vererbung von Tabellen nutzen lässt. Dokumentieren Sie ausführlich Ihre Schritte bei der Übertragung des E/R Modells in die Tabellenstruktur im PostgreSQL

Folgende Quellen sollen für die Datenakquisition genutzt werden:

**TLUG:** http://www.tlug-jena.de/hw/datenladen.html

Die Stationen sollen aus der Ausgangsdate Pegel.csv übernommen werden.

**DWD:** Einladen der DWD Stationen:

https://www.dwd.de/DE/leistungen/klimadatendeutschland/statliste/statlex\_rich.txt?view=nasPublication&nn=16102

DWD: Auswahl von jeweils drei Beispielstationen und den vier Parametern:

http://www.dwd.de/klimadaten (aktuelle Tageswerte)

Überführen Sie diese Daten in Ihre Datenbank. Erstellen Sie für die TLUG Daten Sichten (views), welche die jeweiligen wöchentlichen Durchschnittswerte darstellen. Für die DWD Daten ist eine Übersicht (view) mit dem Jahresmittel für Niederschlag und Temperatur zu erstellen.

Aus Gründen der Nachvollziehbarkeit, sollen sämtliche Änderungen an Stationsdaten nochmals in einer „Historientabelle“ (Station\_log) mitgeführt werden. Erstellen Sie dafür einen Trigger, der sämtliche Änderungen dieser Tabelle in Station\_log protokolliert, zur Kontrolle nehmen Sie selbst einfache Änderungen mit dem update Befehl vor.

1. **Methodisches Vorgehen**

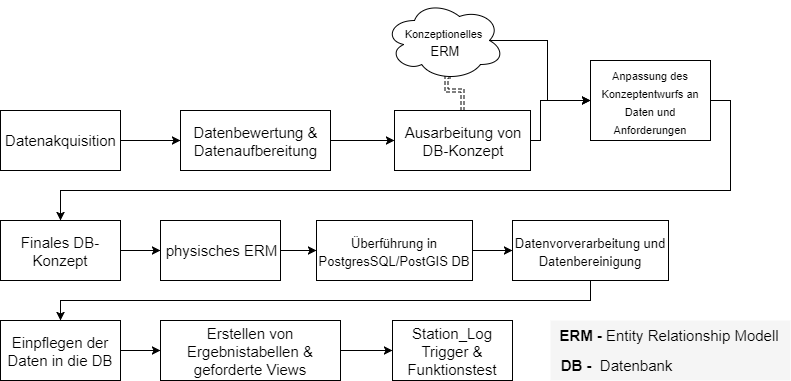
Die wichtigsten Punkte des methodischen Vorgehens im Geodatenbankprojekt, können aus Abbildung 1 entnommen werden. Zu beginn werden die Rohdaten der Klima- und Pegelmessstation von den jeweiligen Einrichtungen (TLUG, DWD) Online bezogen. Nach der Aufbringung aller Rohdaten, welche die Daten der Messstationen und deren Messwerte umfassen, wurde ein erstes konzeptionelles Datenbankmodell erstellt.

Abbildung 1) Methodisches Vorgehen

Im Laufe der Ausarbeitung des Konzeptionellen Datenbankmodells, wurden die Eingangsdaten iterativ auf das Datenbankmodell angepasst, sowie das Datenbankmodell auf die Anforderung der geforderten Datenbank. Im nächsten Schritt wurde das Final Datenbankmodell konkretisiert, da das konzeptionelle Entity-Relationshipmodell nur als grobe Vorlage dienen kann und es für die Erstellung der eigentlichen Datenbank nicht die nötige Detailtiefe besitzt.

Bei der Erstellung des physischen Entity-Relationshipmodells werden alle Entitäten der Datenbank, sowie deren Attribute, Datentypen der Attribute und Relation innerhalb der Datenbank festgelegt. Dabei sind grundlegende Konzepte wie, Normalisierung der Datenbank in zweite und dritte Normalform, die Atomarisierung der Daten innerhalb einer Entität und die Festlegung von Primär- und Fremdschlüssel zu berücksichtigen (Elmasri & Navathe, 2009). Durch die Erstellung des physischen ERM liegt ein genauer Plan zur Erstellung der Datenbank vor welcher nun in PostgreSQL übernommen werden kann.

Bevor die beschafften Messdaten in die Datenbank eingepflegt werden, wurden diese bereinigt und für den Import in die PostgreSQL-Datenbank angepasst. Im letzten Schritt wurden die geforderten Sichten erstellt welche die wöchentlichen und jährlichen Mittelwerte der Messstationen beinhalten. Darüber hinaus wurde ein Datenbanktrigger erstellt, welcher alle Änderungen an der Entität der Messstationen dokumentiert.

1. **Entity Relationship Modell**

Dieses Kapitel beschäftigt sich zu Beginn mit dem ersten Datenbankkonzept (siehe Abbildung 2) und im zweiten Teil mit dem Konkreten Datenbankmodell (ERM, siehe Abbildung 3), welches schließlich auch als physische Datenbank in PostgreSQL/PostGIS umgesetzt wurde. Das konzeptionelle Datenbankmodell (Abb. 2) dient dabei für einen ersten Überblick über möglich Beziehungen und Entitäten innerhalb der Datenbank. Im physischen Datenbankmodell (Abb. 3) werden die Ideen des konzeptionellen Modells konkretisiert und soweit ausgeführt das es als Vorlage zur Umsetzung der Datenbank in PostgresSQL dienen kann.

**3. 1 Konzeptionelles ERM**

Das konzeptionelle Datenbankmodell ist aus Abbildung 2 zu entnehmen. Es wurde sich dafür entschieden das die Tabellen Messstationen und Messdaten im Mittelpunkt stehen sollen, da diese primär in der Datenbank verwaltet werden. Die Tabelle Messstationen wird dabei von den Entitäten Einrichtung, Historientabelle und Meta-Daten verwaltet, dokumentiert und beschrieben. Ein wichtiger Punkt bei der Verknüpfung von Messstationen und Messdaten ist das beliebig viele Messdaten in Relation zu einer Messstation stehen können. In der späteren Umsetzung werden die Tabellen Messdaten und Messstationen nach Klima und Pegel unterschieden und mittels Vererbung an übergeordnete Tabellen weitergegeben.

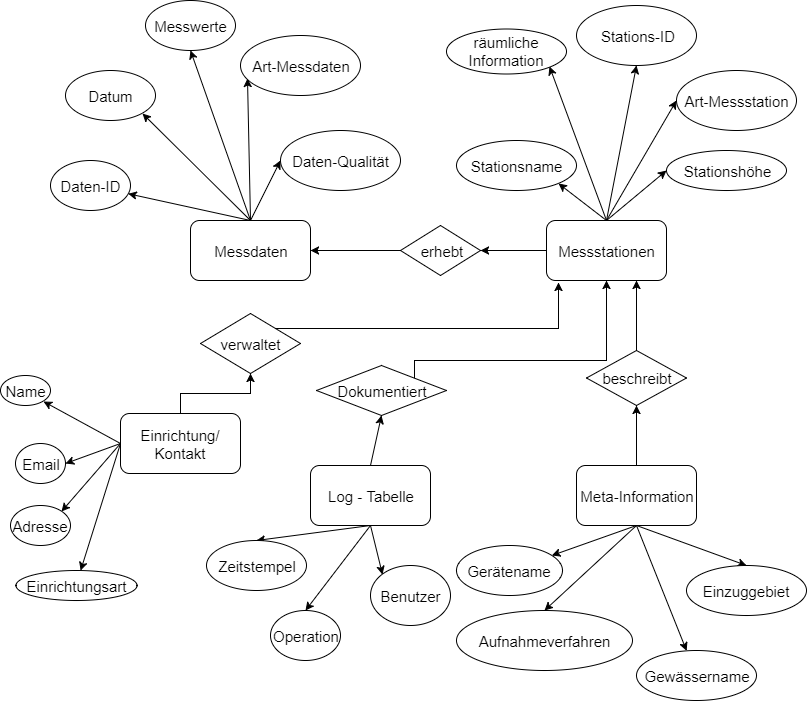
****

Abbildung ) Konzeptionelles Datenbankmodell

**3. 2 Physisches Entity-Relationshipmodell**

Das physische ERM stellt die Finalisierung des Datenbankkonzeptes aus 3.1 dar (siehe Abbildung 3). Dabei sind, wie schon im Konzept erarbeitet, die Entitäten Stationen und Messdaten die wichtigsten Tabellen in der Datenbank. Die Unterscheidung zwischen Klima und Pegel Messstation/-daten wurde über die Art ID (AID) getroffen. Dabei kommt zur Umsetzung der Messdaten auch Vererbung zum Einsatz, in welcher die Tabelle Messdaten die Daten ID (DID), Art ID (AID), Stations\_ID und das Messdatum von den Entitäten Klima\_Messdaten und Pegel\_Messdaten vererbt bekommt. Darüber hinaus wurde die Tabelle VG2500\_BLD hinzugefügt, welche die räumlichen Information der deutschen Bundesländer beinhaltet. Auch wurden die Tabellen für die Ergebnissichten Jahreswerte\_Klima und Wochenwerte\_Pegel hinzugefügt, welche für die Erstellung der Ergebnissichten benötigt werden und so die Wochen- und Jahresmittel archivieren.

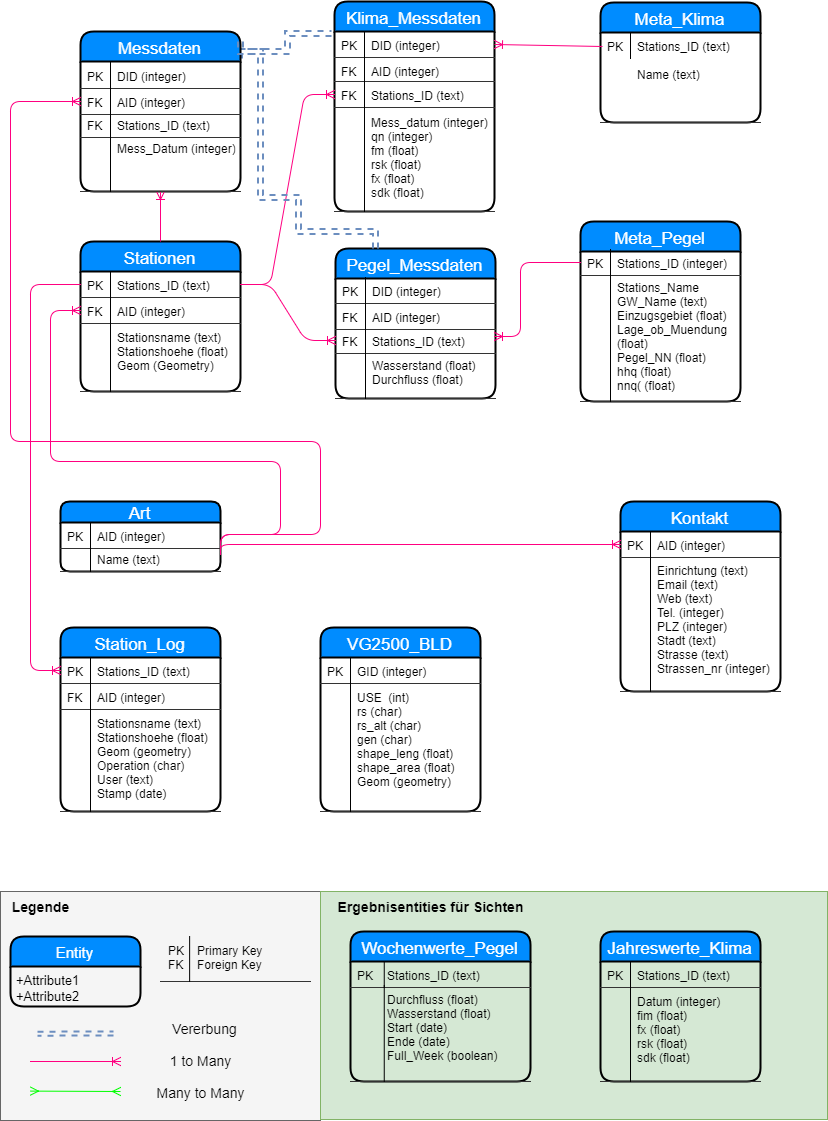
****Bei der Umsetzung der Datenbank wurde auf die Einhaltung der dritten Normalform geachtet, dies meint die Abhängigkeit der Datensätze vom Primär Schlüssel (Elmasri & Navathe, 2009). Dies wurde zu einem großenteil in der Datenbank berücksichtigt, lediglich bei der Umsetzung der Tabellen von Metadaten und Kontakt wurde diese Normalisierung für die Übersichtlichkeit der Datenbank vernachlässigt. Ferner wurden innerhalb der Datenbank nur „one to many“ (1:n) geknüpft, zur Vermeidung von unhandlichen „many to many“ (m:n) Beziehungen.

Abbildung 3) physisches ERM

**4 Überführung in PostgreSQL/PostGIS Messdatenbank**

Für die Erstellung der Datenbank wurde PostgreSQL 10 mit der Erweiterung PostGIS in der Version 2.4.0 genutzt. Darüber hinaus wurde die graphische Benutzerschnittstelle pgAdmin III genutzt um die PostgreSQL Datenbank zu bearbeiten. Alle Programme wurden unter dem Betriebssystem Windows 10 von Microsoft ausgeführt. Das detaillierte ERM diente als Vorlage bei der Datenbankerstellung.

Damit die Erweiterung PostGIS nutzbar wird muss diese explizit im SQL-Code angegeben werden (siehe Abbildung 4).



Abbildung ) PostGIS Extension

Für die Überführung des geforderten Shapefile der Bundesländer in die Datenbank wurde die Anwendung shp2psql genutzt, welche mit Installation der PostGIS-Erweiterung verfügbar ist.

Zur Vermeidung von Datenredundanz/-operationen und für die übersichtliche Gestaltung der Datenbank wurden Vererbungsregeln auf die Tabellen Messdaten, Klima\_Messdaten und Pegel\_Messdaten angewandt (siehe Abbildung 5).



Abbildung ) Vererbung an Messdaten

Durch diese Vererbungsregeln musste kein expliziter „insert“ Befehl an die Tabelle Messdaten übergeben werden, denn die an die Tabelle Messdaten wurden automatische die Variablen DID, Stations\_ID, Mess\_Datum und AID vererbt.

Für die Dokumentation aller Änderungen der Tabelle Stationen war es nötig eine Datenbank-trigger einzufügen, welcher die Änderungen (update - verändern, delete – löschen und insert – einfügen) in der Tabelle Station\_Log dokumentiert (siehe Abbidlung 6). - **QUELLE**

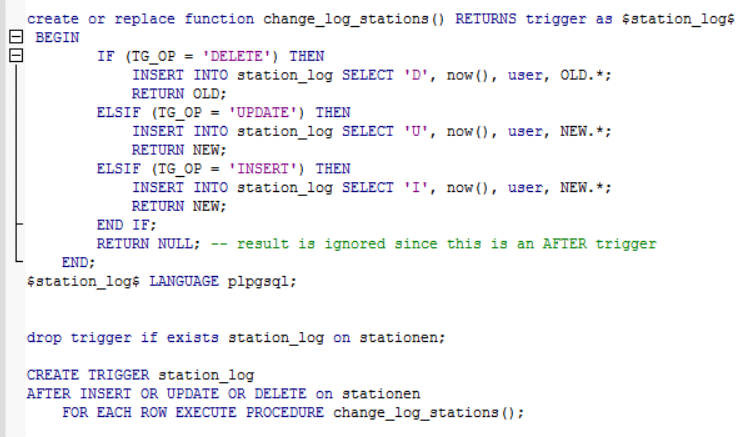


Abbildung 6) Datenbank Trigger

Die Möglichkeit Relation innerhalb der Datenbank zu knüpfen wurde durch Primär- und Fremdschlüssel ermöglicht, wie bereits im ERM dargestellt. Zur Veranschaulichung soll die Beispielabfrage in Abbildung 7 dienen.



Abbildung 7) Beispielabfrage "inner join"

**5 Datenquellen und Vorbereitung**

Für dieses Datenbankprojekt wurden Daten des Deutschenwetterdienstes und Daten der Thüringer Landesanstalt für Umwelt und Geologie genutzt. Aus beiden Datenquellen wurden Messstation und deren Lage bezogen, die Messdaten in Form von Zeitreihen, sowie die Metadaten der Messstationen. Die genutzten Daten teilen sich wie folgt auf:

DWD:

* 3 verschieden Klimamessstationen des Deutschen Wetterdienstes (DWD, 2017)
* Messdaten der 3 Stationen zu Sonnenscheindauer, max. Windgeschwindigkeit, Temperatur und Niederschlag
* 1 Jahr aufgeteilt in 12 Monate

TLUG:

* Pegelmessstationen (Breitgestellt durch Modulleiter)
* Messdaten von 5 verschiedenen Pegelstation zu Wasserstand und Durchfluss
* Zeitraum von 4 Wochen (TLUG, 2017)

Die von DWD und TLUG bezogenen Daten zu Messstationen und Messwerten wurden als Tabstopp getrennte Textdaten heruntergeladen und mit der Programmiersprache R (Version 3.4.2) für den Import in die Datenbank vorbereitet.

**Literatur**

DWD. (2017a). Klimadaten Deutschland. Retrieved from

http://www.dwd.de/DE/leistungen/klimadatendeutschland/klimadatendeutschland.html

DWD. (2017b). Klimastationen Deutschland. Retrieved from

https://www.dwd.de/DE/leistungen/klimadatendeutschland/statliste/statlex\_rich.txt?view

=nasPublication&nn=16102

Moosmeier, J. (2011). Bundesländer Deutschland. Retrieved from

https://www.arcgis.com/home/item.html?id=ae25571c60d94ce5b7fcbf74e27c00e0

Schoening, H.-J. (2013). Tracking changes in PostgreSQL. Retrieved from

http://www.cybertec.at/tracking-changes-in-postgresql/

TULG. (2017). Pegeldaten. Retrieved from http://hnz.thueringen.de/hw2.0/datenladen.html