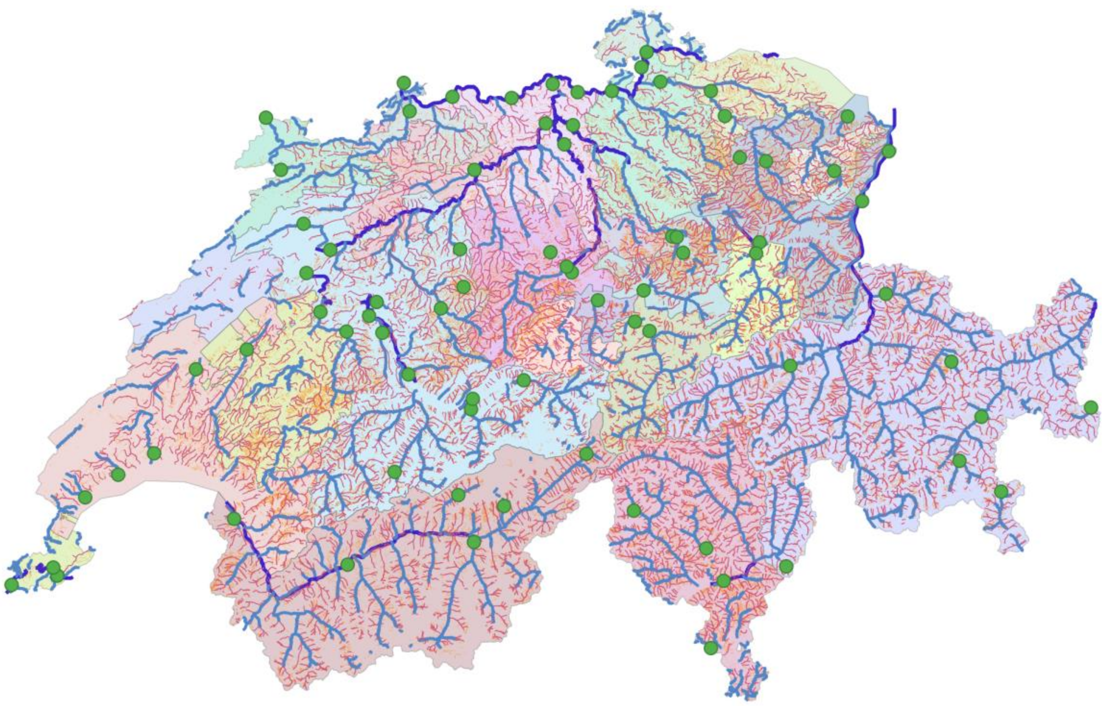
Project: Swiss Water Resources

Project Description

Introduction

Swiss water resources are of major importance concerning ecology and economics. Much of the data being collected about the various resources is used for immediate monitoring of chemical pollution or water flow. Additionally, these data sources serve as a source for various derived information such as building up predictive models and validating possible correlation.

This project organizes data about Swiss water resources for researchers and data analysts. Multiple collaborators will analyze datasets that are centrally hosted and organized by your project team. Geographical information about rivers, stations as well as regional surface information is tightly coupled to measurements and publicly available sensor data. The relevant open data will be processed, stored, and managed in a suitable manner in a spatial database management system – PostgreSQL/PostGIS is used here. Additionally, you will create and store spatial and non-spatial information yourself.



Project Specification

The research project "Swiss Water Resources" needs its data being organized in a database. You and your development team build up the data storing system according to the following requirements.

Part I

Water measurement stations of Switzerland serve as base dataset in this project. The same is true for all Swiss river types represented by classified river segments – also called "river topology". Continuous measurements of all parameters from the station "Basel-Weil" must be available just as the daily means data (discharge and temperature) from the same station. Daily means data from two further stations "Rheinfelden" and "Ilanz" is to be imported, too. Regarding the daily means values: only data of the year 2022 – as provided - is of interest and is part of the database.
Not all attributes of the river type dataset need to be part of the project. In addition to the continuous attributes discharge and slope the following categorized values/attributes shall be implemented: biogeographical regions,

main river membership, elevation type, discharge level, soil type, slope type. A proper dataset of Swiss cantons should be provided in the database for further spatial statistics.

Part II

Scientists working with all previously described data must register in the system/database with their personal information such as name, first name, email address, and registration date. After registering they can do two things. On the one hand they are allowed to write free text comments/suggestions about a measuring station to improve or extend the measuring protocol with new sensors. They have to classify their comments ranging from "not important" over "nice to have" to "super important". The time when the comment was made is stored, too.
On the other hand, scientist can store polygons for their own further use in their spatial analyses. Each geometry must be flagged with a self-assessment of the accuracy for the geometry with five categories from 1 "extremely accurate" to 5 "poor accuracy". A comment can be written to every polygon saved.

Part III

You should generate exemplary data for scientists using the system, storing their information, commenting stations, and materializing their polygons used in their calculations. Insert this data into the database. Finally, you will test whether your database can correctly answer the specified queries. Please note: Queries are also part of the requirement analysis.

Query tasks to be solved

1. Which stations touch a river section?
2. Which river sections are closer than 90 meters to a measure station? Give the main river name (if available) and the measure station name.
3. List the shared river sections between the three cantons Zug, Zürich and Schwyz.
4. Which river sections connect to (=touch) the sections of main river Rhône?
5. What are the yearly averages and standard deviation of the pH rounded to 1 decimal at "Rhein-Weil" since ever?
6. Construct – with SQL – the main river Rhine as one single multiline geometry. How long is the river Rhine in km?
7. How many river sections are completely inside one of your self-created and saved polygons and what is their length and what is the size of the respective polygon in square kilometers? Also include the name of the respective scientist and the quality rating of the polygon. Give a full list with all exemplary polygon records.
8. Do the daily means water temperature of the Weil am Rhein station in 2022 equal to the values from the continuous hourly measurement (aggregated per day)?
9. How many hours per year could we have swimmed in Weil am Rhein at a water temperature of higher than 25C° in the last twenty years?
10. Convert all cantons to geojson format in one result field. Since there are many vertices and a long output, first simplify the cantons geometries by a factor of 500. Test your multi geometry feature collection on https://geojson.io.
11. Precalculated vs. Dynamic. The base river sections dataset has a precalculated biogeographic region as a categorized value. But we can also calculate the same value 'on the fly' with a spatial join. Why do the results from both methods differ? How did the precalculation deal with river sections overlapping more than one biogeographic region?

Feel free to add some more tasks & solutions provided you are done with the tasks mentioned above.

Datasets

Downloadable data & geodata

- Water temperature **monitoring stations** (Federal Office for the Environment FOEN). Official ID, Name of Station, Location: https://data.geo.admin.ch/ch.bafu.hydrologie-wassertemperaturmessstationen/data.zip
- **River typology** for Switzerland: https://data.geo.admin.ch/ch.bafu.typisierung-fliessgewaesser/data.zip
- ❶ **Station 1**: Ilanz (no 2033). Daily means. Year 2022. Date, Discharge, Temperature [two pdf files – annual tables]: https://www.hydrodaten.admin.ch/en/2033.html
- ❷ **Station 2**: Rheinfelden (no 2091). Daily means. Year 2022. Date, Discharge, Temperature [two pdf files – annual tables]: https://www.hydrodaten.admin.ch/en/2091.html

- ❸ **Station 3:** Basel-Weil (no. 2613). Daily means. Year 2022. Date, Temperature only [one pdf file]: https://www.hydrodaten.admin.ch/en/2613.html

  Compiled version of (❶❷❸) containing the extracted mean values from the pdf files [one xlsx File] https://www.geo.uzh.ch/microsite/geo875/scripte/STATION_2033_2091_2613_PDF_DATA.xlsx

- **Station 3:** Basel-Weil (no. 2613). Filter online for 1st of January 2004 until 30th of June 2024 [343'702 records]. FromDateTime (=Start_text), ToDateTime (=Ende_text), Electric Conductivity, Oxygen Content, pH value, Temperature. [one csv file]: https://data.bs.ch/explore/dataset/100046/table/?sort=startzeitpunkt

- Swiss **Biogeographical regions** https://opendata.swiss/de/dataset/biogeographische-regionen-der-schweiz-ch

- Swiss **Cantons** on geo.admin.ch. Visualize the cantonal boundaries. Layer information then shows information about downloading swissBOUNDARIES3D data. On the following link choose downloading the latest release from January 2024 in shape file format. Out of the five resulting geometry datasets, only the files belonging to …KANTONSGEBIET….shp (=cantons) are relevant to us. https://www.swisstopo.admin.ch/de/geodata/landscape/boundaries3d.html

Datasets to be self-created

- Invent some (at least three) exemplary scientists and data they entered to comment stations.
- Put at least five exemplary, self-digitized polygons for each scientist to the database, reflecting their data used for spatial analysis processes.

Instructions

1. Introduction

In Project Geo875 " Swiss Water Resources" you will practically apply the knowledge and skills acquired during the lectures and exercises to the modeling of spatial databases. You will take the role of a research software engineer for your company or research institute. The project goal is to model and implement a spatial database based on a simplified but realistic example from Environment Geography. The duty covers the entire sequence from requirements analysis through the ER-modeling and the derivation of the DB schema up to implementation in a PostGIS spatial database. You should find and process relevant data as well as store and manage the data in a database management system PostgreSQL/PostGIS.

Learning Objectives

- You will deepen your knowledge of the processes and practical skills for planning and implementation of a spatial database by executing all the necessary steps.
- You will strengthen and expand your knowledge of SQL.
- You will collect spatial data and thereby develop knowledge of basic spatial data structures.
- You will deepen your theoretical knowledge of database concepts such as *keys* by linking tables and by inserting data to the database in the correct order.
- You will learn that careful modeling of (spatial) database structures reduces the subsequent implementation effort.
- You will learn to deal with different data formats and load this data into spatial databases.
- You will learn to materialize derived/precalculated results to (derivation-)tables.

2. Tasks

Please study Section 4: Tips and Tricks before you begin.

a) **Requirements Analysis**: The **Project description** presents the results of a requirements analysis for the project. Study this requirements analysis and relevant open data, and identify entity types, relationship types, and possible constraints for the project.

b) **Conceptual Design:** Create the first structure of an **ER model** for the project based on the requirements analysis (= project description). The simplest way to do this is by using MS PowerPoint with *shapes* and *connectors* – find our template file here. The ER-Diagram represents the first milestone in your project and is due in week 3 (Wednesday 20.11.2024). You will present your ER at the feedback meeting that will take place on Friday of the same week (22.11.2024). The following information must be specified on your ER report: **your name**, **group number**, **a date,** and **version** of the ER (e.g., version 1, the ER is going to change!).

c) **Logical design**: Create the logical DB design for this project by translating the ER from Task 2 into a relational model.

d) **Physical Design (Implementation):** Implement the designed database in the PostreSQL/PostGIS spatial database management system. Create the tables according to the relations in the logical design. **Don't forget to add some comments/metadata to each table/column**.

e) **Data Loading**: Load, transfer, and adjust spatial and non-spatial data provided into your project schema in the database. You should also generate exemplary data for the scientist's analysis and insert them into the database.

f) **Database Information Retrieval:** Test your database by executing specified **queries** identified in the project description. If you have modeled everything correctly, your database should now be able to answer the queries, although possibly by using complex SQL statements.

g) **Reporting Progress**: In week 5 (Friday 06.12.2024) you will again report to the course teachers about the progress of your project and receive feedback for the final part of the course.

h) **Project Report**: All steps should be documented in a project report. Wherever possible use formal specifications used in the lecture (ER, relational model, SQL-listings). The project report serves as your own documentation as well as the source for the evaluation by the course teachers. Find a good balance between work log, SQL-listings and interpretation and discussion of your work steps. We are interested not only in *what* you have done, but also *why* you have done something.

3. Data

- You can find a list of relevant datasets in the **Project Description** part in this document. Please note that these datasets might be in different data formats (e.g., CSV, shapefile, geopackage, Excel);
- Invent at least three users/scientist each of them storing at least five polygons;
- If you have enough time in your project and done all the tasks defined in the project description, think about
  - ☺ modelling and incorporating other open datasets
  - ☺ inventing more advanced query tasks and their solution with SQL

4. Tips and Tricks

- Swiss government and cantons provide many open data datasets. Please note that these datasets might contain many additional attributes. You might only want to store the "important" ones. We tried to mention the relevant ones in the **Project Description** part in this document.
- Watch out for denormalized data or hidden categorized values in some files and spatial datasets! Additionally, some attributes seem to be perfect as meaningful primary key attributes - but check the data to be sure: If there are no natural primary key attributes, create your own identification column (=surrogate key column).
- When you invent the scientist's exemplary data, you might want to use QGIS to draw polygons and store their geometry into the database. Inserting and updating points and even simple lines and polygons can be an easy job with pure SQL, too. This should be your choice for all non-spatial tables and to deal with all kinds of attributes.
- Use the CH1903+ LV95 (EPSG: 2056) projection in your QGIS project. Transform your spatial datasets in the database to the correct geometry type and the CH1903+ LV95 projection. The latter you preferably do even before importing your data to the database.
- The **online project website** https://www.geo.uzh.ch/microsite/geo875/project.html has more tips on **data analysis** for ER modelling/DB design, **data importing** and writing **the report**.
- The FAQ page is more database and SQL centric: https://www.geo.uzh.ch/microsite/geo875/faq.html.

5. Timetable

Wed 20.11.2024		Handover of ER diagram: digital PDF version to zhiyong.zhou@geo.uzh.ch
Fri	22.11.2024	Feedback ER discussion with the group
Fri	06.12.2024	Discussion with the group about the current state of the project
Fri	20.12.2024 11.59 AM	Handover of the final project report: digital PDF version to zhiyong.zhou@geo.uzh.ch (see the communication rules on the geo875 microsite).