Data format for integrated indexing on spatial data

**Keywords:** Scientific data management, spatial indexing, data format

**Problem:** Scientists in many disciplines generate massive amounts of data. The amounts of data produce are already so big that they can barely be managed. In order to improve the data analysis capabilities of data management tools, scientists have created data formats which along with the original data store also metadata and indexing information (e.g., MBRs of an R-Tree index). Data formats following such specification is “shapefiles”.

**Project:** The goal of this project is to analyze the existing data formats that store additional information on the side and recognize the possible disadvantages of each approach and understand why such data formats are not more widely spread. In addition, based on the recognized disadvantages, design a new data format that is generic enough to be used by multiple existing indexing structures while keeping the storage overheads low.

**Plan:**
1. Recognize the different data formats and implement a reading/querying mechanism for them into PostgreSQL.
2. Understand the disadvantages of the different data formats.
3. Design a prototype that will keep low storage overhead while store sufficient information to improve query performance
4. Run experiments to

**Supervisor:** Prof. Anastasia Ailamaki, anastasia.ailamaki@epfl.ch

**Responsible collaborator(s):** Pavlovic Mirjana, mirjana.pavlovic@epfl.ch
Matthaios Olma, matthaios.olma@epfl.ch
Manos Karpathiotakis, manos.karpathiotakis@epfl.ch

**Duration:** 2-3 monhts