# Was sind Offene Daten

Open Data bezeichnet frei zugängliche und nutzbare Daten, die von öffentlichen oder privaten Organisationen bereitgestellt werden, um Transparenz, Innovation und Beteiligung zu fördern. Diese Daten sind in der Regel maschinenlesbar und können ohne Einschränkungen weiterverwendet und verbreitet werden.

# KVB Offene Daten

Die [offenen Daten der KVB](https://kvb.koeln/service/open_data.html) liegen im JSON-Format vor. Grobe Inhaltserklärung der Dateien.

|  |  |
| --- | --- |
| haltestellen.json | Haltestellen, mehrere pro Haltestellenbereich möglich. |
| haltestellenbereiche.json | Haltestellenbereiche |
| haltestellenbereichemitdfi.json | Haltestellenbereiche mit digitaler Fahrgastinformationsanzeige |
| verkaufsorte.json | Fahrkartenverkaufstellen, den Haltestellenbereichen zugeordnet. |
| fahrtreppen.json | Fahrtreppen, den Haltestellenbereichen zugeordnet. |
| fahrtreppenstoerungen.json | Störungen an Fahrtreppen mit Zeitstempel |
| aufzuege.json | Aufzüge, den Haltestellenbereichen zugeordnet. |
| aufzugstoerungen.json | Störungen an Aufzügen mit Zeitstempel |

# Die Aufgaben

# Daten erforschen

Wie hängen die Daten zusammen? Gibt es einheitliche Identifizierungsmerkmale, um Daten anzureichern? Welche Werte liegen hinter den Schlüssel der einzelnen „Dictionaries“. Welche der Dateien hat die meisten Daten, welche eignet sich als Datenbasis? In welcher Form liegen die Koordinaten vor, welcher Wert steht für was? Wie sind die verschiedenen Daten verbunden (Haltenstellenbereich, kurzname, Betriebsbereich)? Würde es Sinn machen Informationen die in verschiedenen Dateien liegen in eine der anderen einzupflegen?

# Teams bilden

Bildet 2er, 3er oder 4er Teams und besprecht den Ablauf des Projekts. Protokolliert eure Planung und Aufteilung.

# Planung einer Funktionsbibliothek

Teil 1 des Projekts ist die Planung und Entwicklung einer Bibliothek die Hilfsfunktionen für den Zugriff auf die Daten. Findet allgemeine Hilfsfunktionen, die die Daten einlesen oder aktualisieren. Plant Funktionen die rudimentäre Aussagen über die Datensätze machen können. Konsultiert dazu auch die Auswertungen, die im Teil 2 des Projekts angefertigt werden müssen. Schaut ob anders sortierte Dictionaries bei der weiteren Arbeit helfen könnten.

Legt die Bibliothek als Modul der einfachsten Form an. Also als eine Datei im gleichen Ordner wie das Hauptprogramm. Nutzt die Möglichkeiten der Dokumentation durch Docstrings und pydoc. Die Bibliothek wird im späteren Verlauf des Kurses nochmal genutzt.

# Auswertung der Daten

Beantwortet unter Zuhilfenahme der Funktionen in der Bibliothek die folgenden Fragen. Schreibt für jede der Fragen eine eigene Funktion. Sollte über die Funktionen redundanter Code auftauchen, überlegt ob und wie man diesen auch in die Bibliothek auslagern könnte. Erweitert die Bibliothek bei Bedarf. Die Aufgaben sind in 4 Schwierigkeitsgrade unterteilt. Sie bauen aufeinander auf.

Schwierigkeitsgrad - leicht

* Wie viele Linien (Busse und Bahnen) und wie viele Haltestellenbereiche gibt es insgesamt?
* Welcher Haltestellenbereich hat die meisten Haltestellen und an welchem Treffen sich die meisten Linien?
* Wieviel Haltestellenbereiche haben Fahrtreppen, Aufzüge, beides oder keines von beidem?
* Wieviel Haltestellenbereiche decken jeweils die Betriebsbereich „BUS“, „STRAB“ oder beides ab?
* Bestimme den nördlichsten, westlichsten, südlichsten und östlichsten Haltstellenbereich. Wie kann man die Position eines Haltestellenbereichs berechnen?

Schwierigkeitsgrad – mittel

* Welche beiden Linien haben die größte Anzahl gemeinsamer Stationen?
* Welche Linie ist an all ihren Haltestellen prozentual am besten mit digitalen Fahrgastinformationsanzeigen ausgestattet? Das gleiche auch für Fahrtreppen und Aufzüge.
* Der Kölner weiß das eine zweistellige Nummer eine Straßenbahn und eine dreistellige Nummer ein Bus ist. Aber könnte man diese Information auch aus den Daten ziehen? Direkt angegeben ist es nicht.
* Liste alle Straßenbahnlinien jeweils mit allen Straßenbahnlinien in die man auf ihrem Weg direkt umsteigen kann auf.
* Haben alle Straßenbahnlinien mindestens ein Kundencenter an einer ihrer Haltestellen? Und von welchen Linien ohne Kundencenter kann man mit einmal Umsteigen in eine andere Linie zu einem Kundencenter gelangen.

Schwierigkeitsgrad - schwer (Die Erde ist eine Scheibe – behandele Längen und Breitengrad als X,Y Koordinaten)

* Suche für eine eingegebene Position:
  + Die nächste Straßenbahnhaltestelle
  + Die nächste Straßenbahnhaltestelle mit Fahrtreppen (Störung?)
  + Die nächste Straßenbahnhaltestelle mit Aufzug (Störung?)
* Leider gibt es in den Daten keine direkten Informationen in welcher Reihenfolge die einzelnen Linien die Haltestellen abfahren. Könnte man das anhand der Vorgabe einer der beiden Endhaltestellen und den „Entfernungen“ der Haltestellen durch ihre Koordinaten berechnen. Wie akkurat wäre diese Berechnung? Gebt dazu einer der Strecken von Hand ein und vergleicht. Entwickelt vorher die Datenstruktur, in der alle Haltestellen aller Linien in ihrer Reihenfolge gespeichert werden können.

Schwierigkeitsgrad – unmenschlich (Die Erde ist doch rund, um genau zu sein, eine Kugel)

* Entwickelt eine Funktion die die Entfernung zweier Koordinatenpaare in Meter/Kilometer berechnen kann. Stichwort :[„Orthodrome“](https://de.wikipedia.org/wiki/Orthodrome). Überarbeitet die Funktionen der Bibliothek die mit Entfernungen arbeiten, das mit Meter oder Kilometer arbeiten. Nutzt dazu die Erde als idealisierte Kugel mit einem Radius von 6371 km. Lasst euch dabei von Google oder dem LLM eures Vertrauens helfen.

Viel Erfolg!