Data Management – Open Data Platforms

Following is **Part 2** of the **Portfolio Examination** of the module **Data Management** in **Winter Semester 2023/2024**. Please **upload** your **documents** and **program files** to the corresponding course in **Moodle** Learn **latest until 21.12.2023 at 23:59**.

Open-Data Schleswig-Holstein

Please use the portal "Open-Data Schleswig-Holstein" (https://opendata.schleswig-holstein.de) to develop Python programs to demonstrate the use of data schemas, RDF files, SPARQL, and CKAN due to the following guidelines:

1. Select actual data sets which are public domain ("gemeinfrei") and support linked data and are available as RDF files. Download the newest versions of the RDF files for data processing.

2. Develop a Python program to demonstrate the use of RDF files by using the Python package RDFLib. Examine the RDF file for schools ("Schulen"). Which data schema is used due to 'schema.org'? How does the data schema look like in JSON-LD format?

3. Examine the RDF file for police stations ("Polizeidienststellen"). Use SPARQL to query all police stations located in "Kiel". Provide another example query of your choice.

4. Use the CKAN Action API endpoint of the portal "Open-Data Schleswig-Holstein" (https://opendata.schleswig-holstein.de/api/action) to explore the organizations, packages, groups, resources, licenses, and other data resources in a Python program. Demonstrate the use of the CKAN Action API with examples. How can the data sets of the portal be found and accessed via CKAN?

The **deliveries** are the Python program file(s) or Jupyter notebook(s) including the JSON response(s) and the answers to the above questions.

Open Data Platform

Please prepare a **presentation of 5 minutes** of an Open Data Platform of your **confirmed choice** to be held **in person** on **18.12.2023 at 8:30** in the lecture at university. Consider the following guidelines for the content of your presentation:

1. What is the Open Data Platform about and who is the organization behind?

2. Which kind of data sets are stored? Which data formats are supported?

3. Which catalogue system is used? Which kind of API is provided?

4. Which open data quality is available due to the 5-star open data plan?

The **deliveries** are the presentation (as a PDF file).